NAVPERS 93400A-4

FUNDAMENTALS OF ELECTRONICS

VOLUME 4

RECEIVER CIRCUIT APPLICATIONS

Bureau of Naval Personnel

PREFACE

This book is part of an nine-volume set entitled "Fundamentals of Electronics". The nine volumes include:

Volume 1a-NavPers 93400A-1a, Basic Electricity, Direct Current
Volume 1b-NavPers 93400A-1b, Basic Electricity, Alternating Current
Volume 2 - NavPers 93400A-2, Power Supplies and Amplifiers
Volume 3 - NavPers 93400A-3, Transmitter Circuit Applications
Volume 4 - NavPers 93400A-4, Receiver Circuit Applications
Volume 5 - NavPers 93400-5, Oscilloscope Circuit Applications
Volume 6 - NavPers 93400-6, Microwave Circuit Applications
Volume 7 - NavPers 93400-7, Electromagnetic Circuits and Devices
Volume 8 - NavPers 93400-8, Tables and Master Index

If you are becoming acquainted with electricity or electronics for the first time, study volumes one through seven in their numerical sequence. If you have a background equivalent to the information contained in volumes one and two, you are prepared to study the material contained in any of the remaining volumes. A master index for all volumes is included in volume eight. Volume eight also contains technical and mathematical tables that are useful in the study of the other volumes.

A question (or questions) follows each group of paragraphs. The questions are designed to determine if you understand the immediately preceding information. As you study, write out your answer to each question on a sheet of paper. If you have difficulty in phrasing an answer, restudy the applicable paragraphs. Do not advance to the next block of paragraphs until you are satisfied that you have written a correct answer.

When you have completed study of the text matter and written satisfactory answers to all questions on two facing pages of the book, compare your answers with those at the top of the next even-numbered page. If the answers match, you may continue your study with reasonable assurance that you have understood and can apply the material you have studied. Whenever your answers are incorrect, restudy the applicable material to determine why the book answer is correct and yours is not. If you make an honest effort to follow these instructions, you will have achieved the maximum learning benefits from each study assignment.

Follow the directions of your instructor in answering the review questions included at the end of each chapter.

TABLE OF CONTENTS

CHAPTER 29

SOLID STATE POWER SUPPLIES

Most of the people employed in the field of electronics are aware of the growing popularity and importance of the transistor. Yet few people, not directly involved with them, really understand what a transistor is or its basic operation. The transistor is only one phase of the vast field of electronics that falls under the heading of SOLID STATE or SEMICONDUCTOR electronics.

The purpose of this chapter is to introduce solid state physics and semiconductor theory as a basis for the study of transistors. To understand semiconductor theory a knowledge of atomic structure is needed. While Chapter 1 explained atomic structure, a review is in order at this time and will be given with emphasis on specific areas.

SOLID STATE PHYSICS

29-1. The Structure of Matter

Matter is defined as any substance that has mass and occupies space. If some substance, such as a quantity of water, were to be divided and then divided again and again until a particle was obtained which could not be further divided and still be called water, a MOLECULE of water would be the result. In other words, the smallest particle that matter can be divided into and still retain its original identity is a molecule.

A molecule is composed of one or more ELEMENTS. An element can be defined as a substance which cannot be decomposed by ordinary chemical changes, nor made by chemical union of a number of substances. For example, if a molecule of water was divided the water would cease to exist and in its place would be the elements hydrogen and oxygen. An element can be subdivided into ATOMS. The atom is the smallest part of an element that can participate in ordinary chemical changes. Atoms of different elements differ in their average mass, but atoms of the same element are of the same average mass. The atom may be considered to be the smallest part that retains its identity as part of the element from which it is divided.

It was once believed that the atom was the smallest particle of matter; it has since been determined that the atom itself can be subdivided into still smaller, or SUBATOMIC particles. A single atom is similar to the solar system in that there is a central body, called a NUCLEUS, about which a number of smaller particles, called ELECTRONS, move in approximately elliptical orbits.

Careful study and research has disclosed that each electron in an atom has a charge of electricity that is identical with the charge on any of the other electrons. The charge that is associated with an electron is the smallest electrical charge that has been discovered. For this reason it is called an ELEMENTAL CHARGE. It was arbitrarily assigned a NEGATIVE polarity, thus, an electron is said to exhibit a negative charge.

If we would magnify the nucleus enough, it would be seen to be composed of several different kinds of particles, two of which are the proton and the neutron. The proton has an elemental positive charge, whereas the neutron is an uncharged particle. The positive charge of a proton is equal in magnitude to the negative charge of an electron, but opposite in nature. The atom is said to be electrically neutral when there is an equal number of positive and negative charges, that is, when the number of protons in the nucleus is equal to the number of electrons in orbit.

29-2. Atomic Structure

The structure of an atom is best explained by analyzing the simplest of all atoms, the hydrogen atom. The hydrogen atom is composed of a nucleus containing one proton and a single planetary electron. The elliptical path followed by an orbiting electron makes calculations difficult for even a single planetary electron and virtually impossible for complex atoms. Niels Bohr, a Danish physicist, developed a concept which permitted approximate calculations. Basically, Bohr's concept consisted of averaging the elliptical paths of the orbiting electrons into circular paths with fixed distances from the nucleus. As the electron revolves around the nucleus it is held in this orbit by two counteracting forces. One of these forces is called the CENTRIFUGAL FORCE, and is the force which tends to cause the electron to fly outward as it travels around its circular orbit. This is the same force which causes a car to roll off a high-

way when rounding a curve at too high a speed. The second force acting on the electron is CENTRIPETAL FORCE. This force tends to pull the electron in toward the nucleus and is provided by the mutual attraction between the positive nucleus and the negative electron. At some given radius the two forces will exactly balance each other providing a stable path for the electron.

Q1. If centrifugal force is trying to force a car off the highway on a turn, what provides the equivalent centripetal force to prevent this?

29-3. Energy Levels

Since the electron in the hydrogen atom has both mass and motion it contains two types of energy. By virtue of its motion the electron contains KINETIC energy. Due to its position it also contains POTENTIAL energy. The total energy contained by the electron (kinetic plus potential) is the factor which determines the radius of the electron orbit. As the distance of the orbit from the nucleus increases, the energy required to maintain an electron in that orbit increases. In other words, the highest energy orbits are those farthest from the nucleus.

Each atom has specific energy levels at which an electron is permitted to establish an orbit (these levels vary from the atoms of one element to the atoms of another element but all the atoms of a single element have the same energy levels). These permissible areas for establishing orbits are called SHELLS. These shells are designated by either number or letter as shown in Figure 29-1. The lowest possible energy level at which an electron may establish an orbit is the first or K shell. The shells or permissible energy levels of the atom EXIST WHETHER THEY ARE OCCUPIED BY AN ELECTRON OR NOT (the same as houses exist whether people live in them or not). Therefore, even though an atom only has one electron in the K shell the other shells still exist. Application of energy to the electrons in the K shell (through collision with a photon of energy, for example) will cause the electron to jump to one of the higher permissible energy levels. An electron cannot exist in the space between permissible energy levels. This indicates that the electron will not except a photon of energy unless it contains enough energy to elevate the electron to one of the allowed energy levels.

Once the electron has been elevated to an orbit higher than its lowest possible energy level the atom is said to be in an EXCITED state.

The manner in which the energy levels for the various shells are established is somewhat complicated and is part of a science known as

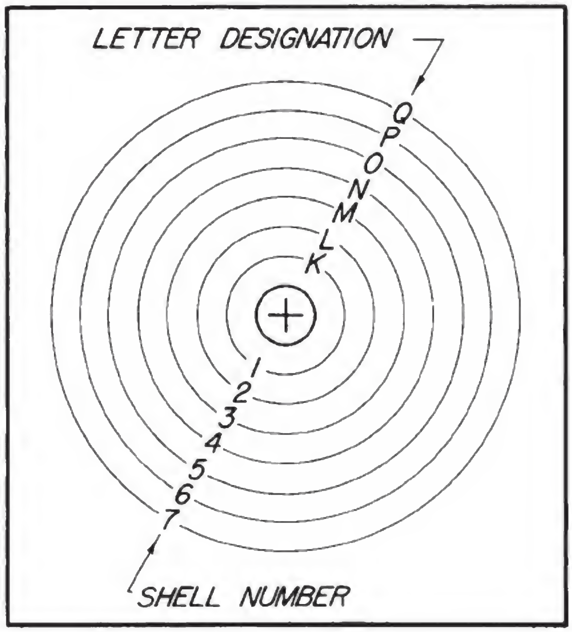

Figure 29-1 - Shell designation.

QUANTUM MECHANICS. In order to determine the orbit of an electron, four numbers called quantum numbers must be known. For an approximate description of an orbit, such as will be used in this chapter only the PRINCIPAL quantum number will be used. The principal quantum number specifies the shell; 1st (K), 2nd (L), etc. It has been determined, by the science of quantum mechanics, that each shell may contain a certain maximum number of orbital electrons. The maximum number of electrons permitted in any shell may be determined by taking two times the principal quantum number squared. Stated mathematically this is:

$$2 \times n^2$$

where n equals the principal quantum number (shell number).

Example. The maximum number of electrons permitted in the first (K) shell would be:

$$2 \times 1^2 = 2 \text{ electrons}$$

The maximum number of electrons permitted in the second (L) shell would be:

$$2 \times 2^2 = 8 \text{ electrons}$$

Table 29-1 lists the maximum number of orbital electrons permitted in each shell. It should be noted that while these are the maximum numbers, and no shell can contain more elec-

trons, it is possible for a shell to contain less or even no electrons.

According to an important theory, known as the PAULI EXCLUSION PRINCIPLE, no two electrons can be described by the same set of quantum numbers in a given closed system, such as an atom. All this means is that no two electrons in a given atom will ever share the exact same orbit.

Shell No.	Letter Designator	Maximum number of electrons	
1	K	2	2×1^2
2	L	8	2×2^2
3	M	18	2×3^2
4	N	32	2×4^2
5	O	50	2×5^2
6	P	Not discovered yet.	
7	Q		

Table 29-1 - Maximum possible occupancy of shells.

The shell is defined as an area of permissible energy levels where electrons may establish orbits. The shells are separated by energy gaps, called FORBIDDEN REGIONS, in which no electrons may establish orbits.

Q2. Would an electron in the N shell contain more or less energy than an electron in the number 2 shell?

29-4. Subshells

Since no two electrons in an atom may share the exact same orbit there must be more than one orbit available in each shell. Each shell is subdivided into SUBSHELLS. There are n subshells in each shell (where n equals the principal quantum number). In other words the first shell will contain one subshell (n=1), the second sub shell will contain two subshells (n=2), etc. The subshells are specified by letter designators. The sequence, starting with the lowest energy level subshell, being s, p, d, f, g, h, i, etc. Although the letters indicate more than four subshells no natural element has been discovered in which any but the first four subshells are used. The letters of the subshells are derived from the science of spectroscopy and have no significant meaning in the field of electronics other than that of designation. While the shells are separated by relatively large energy gaps the difference in energy levels between the subshells (of the same shell) are so small that they are considered to exist one on top of the other. The subshells are also limited as to the maximum number of electrons that they may contain. Table 29-2 lists the maximum number of electrons allowed to occupy each subshell.

letter designator	max. no. of electrons
s	2
p	6
d	10
f	14

Table 29-2 - Maximum possible occupancy of subshells.

Figure 29-2 illustrates the position of the shells and subshells, in relation to the nucleus, by use of an energy level diagram. It can be seen that all shells contain an s subshell and, from the second shell out, a p subshell, etc. Therefore, to differentiate between subshell electrons of the various shells the subshell designator is preceded by the shell number. For instance, "electrons in the 4d subshell exist at a higher energy level than those in the 3d subshell."

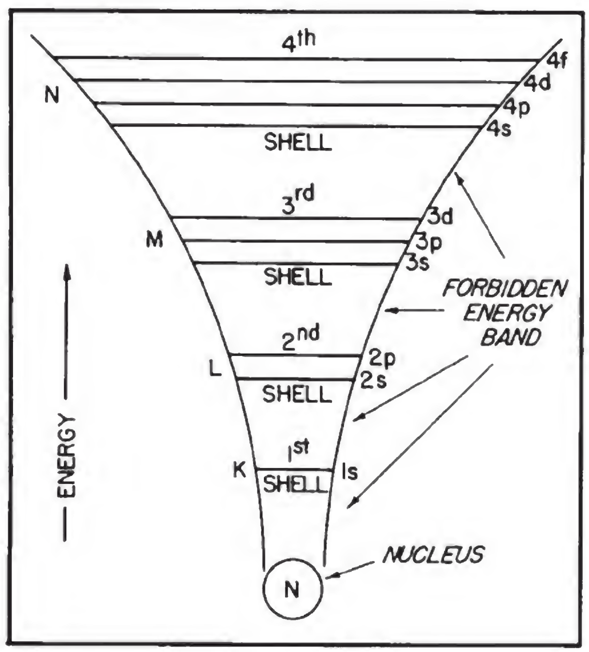

Figure 29-2 - Energy level diagram of shells and subshells.

A1. The friction of the tires on the road surface. When the centrifugal force overcomes the centripetal force the car will skid.

A2. More energy. Higher energy shells are farther from the nucleus.

It must be stressed that Figure 29-2 is a representative diagram only. It was stated that the subshell energy levels, of a given shell, are so close together that they form practically a continuous band of energy levels. For example the 2s and 2p subshells are very close together, while the 2p and 3s subshells are relatively far apart. It must also be remembered that each subshell consists of an energy level for every electron which occupies it. Therefore, the 3s subshell will contain two energy levels, the 3p subshell will contain six energy levels, and the 3d subshell will contain ten energy levels. The 3rd shell (M) consists of 18 energy levels piled one on top of the other and when the M shell is filled an electron will occupy each separate level.

Q3. There is no known natural element with a filled O shell, but if there were, how many subshells would the O shell contain?

Q4. If an electron in the first subshell of the N shell is designated as a 4s electron, how is an electron in the forbidden region designated?

29-5. Valence Electrons

The ATOMIC NUMBER of an element can be used to determine the total number of electrons in the atom, and from this the number of electrons in the outer shell. The atomic number of hydrogen is one, indicating that hydrogen has one orbital electron. The one orbital electron will be in the first or K shell. Since hydrogen only has one electron the K shell is also the outer-most shell. The atomic number of silicon is 14, indicating 14 orbital electrons. Since orbital electrons occupy the lower energy levels first they will be distributed in the silicon atom as follows:

K shell filled = 2 electrons
L shell filled = 8 electrons

The four remaining electrons will establish orbits in the M shell with two of them in the 3s subshell and two of them in the 3p subshell. In the case of silicon the M shell is the outer-most shell with the 3p electrons being the outer-most electrons.

The electrons in the outermost shell are the ones which enter into chemical or electrical combinations with other atoms. These electrons are called VALENCE ELECTRONS and the outermost shell is called the VALENCE SHELL.

When the outermost shell of an atom contains eight electrons the atom is considered to be very stable and will not attempt to gain or lose an electron. For this reason no element will contain more than eight electrons in its outer shell. For example, the M shell can contain 18 electrons when completely filled; however, if it is the outermost shell, it will contain no more than eight electrons.

The VALENCE NUMBER indicates the combining ability of an element. In other words, the valence number indicates the number of bonds that one atom of an element can form with another atom of the same or different elements.

Q5. The atomic number of germanium is 32 and its valence number is four. Which shell is the valence shell? How many valence electrons does it contain and how many bonds can a germanium atom form?

29-6. Energy Bands

When atoms are brought close together, as in a crystal, there is an interaction between the individual energy levels of the atom. Figure 29-3 illustrates the energy level diagrams of two silicon atoms that are separated by a distance large enough to prevent the valence shells from overlapping.

NOTE: The interatomic distances of atoms in a crystal are fixed, according to the type of crystal, and can not be varied. This approach is taken here merely for purposes of illustrating the interaction of energy levels in a crystal.

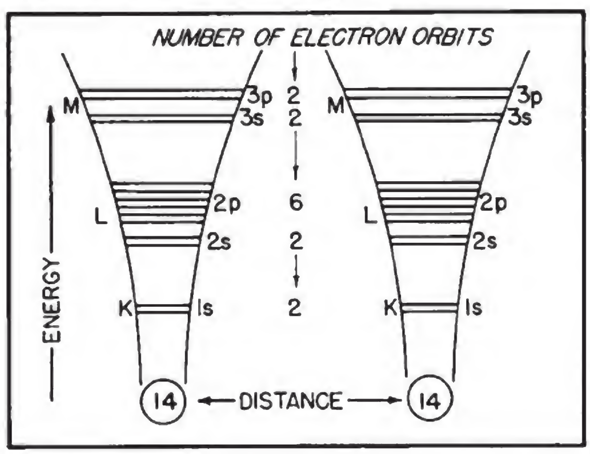

Figure 29-3 - Isolated silicon atoms (energy level diagrams).

The valence shells of each atom have the same energy levels (Figure 29-3) because there is no interaction. Figure 29-4 illustrates the result when the atoms are moved together so that the valence shells just begin to overlap. Since no two electrons may exist at the exact same energy level in a closed system, the energy levels of the valence shell split. The exact way in which this splitting is accomplished is determined by quantum mechanical methods and will not be discussed in this text. Only the outermost shell of each atom has been effected, the inner shells, K and L, retain the energy levels of the individual atom. As more and more atoms are brought together, to form the crystal, the valence shell energy levels continue to split until they form a solid band of discrete permissible energy levels. This band of energy levels is called the VALENCE BAND.

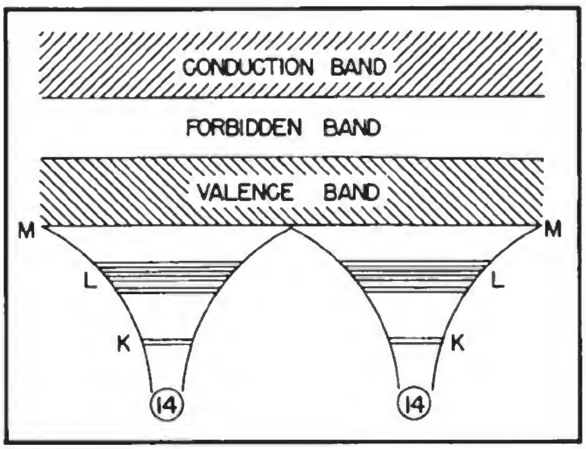

Figure 29-4 - Energy level interaction of silicon atoms.

It was stated that energy levels exist whether they are occupied by electrons or not. These unoccupied energy levels also split when brought together. Therefore, even though the N shells of the silicon atoms are unoccupied they will split, when brought together in the crystal, forming a band of available energy levels. This band of available energy levels is called the CONDUCTION BAND and is shown in Figure 29-4 above the valence band. The conduction band and valence band are separated by the forbidden region which originally separated the M and N shells.

In the study of semiconductor electronics, only the valence band, conduction band, and the forbidden region which separates them is of interest. Since the energy levels below the valence band are not changed, and are of no concern, they will be omitted from the energy level diagrams and only the top bands will be used.

Q6. Which shells of the germanium atoms form the valence band in a germanium crystal and which shells form the conduction band?

29-7. Conductors, Semiconductors and Insulators

The valence band contains electrons that are under the influence of the nucleus. However, the application of the proper amount of energy will cause the electrons to be elevated to the conduction band. The applied energy must be a discrete amount, that is, enough to move the electron all the way across the forbidden region. The valence electrons will never go half way and stop. When an electron reaches the conduction band it is considered to be free from the influence of the parent atom and able to move about in the crystal. It is therefore called a FREE ELECTRON. The forbidden band (in Figure 29-4) is a representation of the amount of energy required to elevate the electron to the conduction band.

Substances that permit the free motion of a large number of electrons are called CONDUCTORS. Copper wire is considered a good conductor because it has many free electrons. Electrical energy is transferred through conductors by means of the movement of free electrons that migrate from atom to atom inside the conductor. The greater the number of electrons that can be made to move in a material under the application of a given force the better are the conductive qualities of that material. A good conductor is said to have a low opposition or low resistance to current (electron) flow.

In contrast to good conductors, some substances such as rubber, glass, and dry wood have very few free electrons. In these materials large amounts of energy must be expended in order to break the electrons loose from the influence of the nucleus. Substances containing very few free electrons are called POOR CONDUCTORS, NONCONDUCTORS, or INSULATORS. Actually, there is no sharp dividing line between conductors and insulators, since electron motion is known to exist to some extent in all matter.

It is the degree of difficulty in dislodging the planetary electrons from the outermost shell of the atom that determines whether the element is a conductor, an insulator, or a semiconductor. The height of the forbidden band, or energy gap, is an indication of the amount of energy required to free a valence electron. The gap energy is measured in electron volts (the electron volt is explained in section 16-2). In a copper crystal the interatomic distances between the atoms are such that, due to interaction, the energy levels of the conduction band overlap those of the valence band. Since no forbidden band exists in a

A3. Five subshells. There are n subshells in each shell and for the O shell n=5.

A4. An electron in the forbidden region is not designated because no electrons may exist in this area.

A5. The N shell is the valence shell. It contains four valence electrons and can form four bonds.

A6. The N shells form the valence band and the energy levels of the O shell form the conduction band.

copper crystal the gap energy is zero electron volts (0 ev). This indicates that at normal temperatures all the valence electrons will be elevated to the conduction band, for this reason copper is a good conductor. On the other hand the interatomic distances of the atoms in a diamond cause the conduction band and valence band energy levels to be separated by a large (7 ev) forbidden band. Thus, a large amount of energy is required to break a valence bond in a diamond crystal, making diamond a very good insulator. The gap energy of a semiconductor material will lie somewhere between the extremes of the conductor and the insulator. For instance, at room temperature the approximate energy gap of germanium is 0.7 ev. Therefore, the energy required to elevate an electron to the conduction band of a semiconductor material is less than that required for an insulator but more than that required for a conductor. Figure 29-5 shows energy level diagrams for a conductor, semiconductor, and an insulator. A comparison of the forbidden energy gaps gives an indication of the amount of energy necessary to elevate an electron to the conduction band in each.

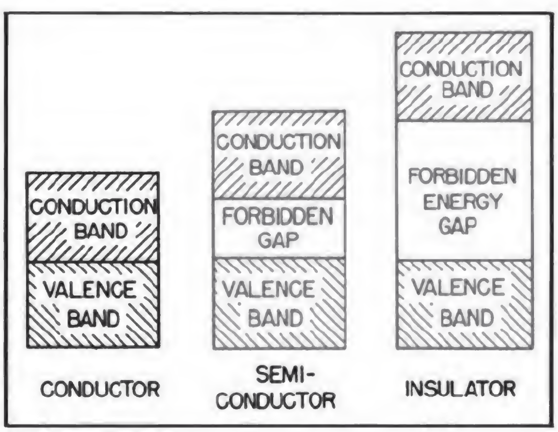

Figure 29-5 - Energy diagrams for conductor, semiconductor and insulator.

The energy diagram for the insulator shows a very wide energy gap. The energy gap of the conductor is nonexistant, indicating that very little if any energy is required to elevate an electron to the conduction band.

Q7. The gap energy of silicon at room temperature is approximately 1.1 ev. Is silicon a conductor, semiconductor, or insulator?

29-8. Covalent Bonding

The valence of an element is a measure of its ability to combine with other elements. It depends on the number of planetary electrons in the outer shell of the atom. For example, hydrogen has a valence of one and will combine readily with other atoms. The reason is that the outer shell of the hydrogen atom has only one planetary electron and requires one more electron to be complete. Oxygen has a valence of 2 because the outer shell has only 6 planetary electrons; it needs 2 more to make up the full complement of 8.

Two atoms of hydrogen combine with one atom of oxygen to form a molecule of water, Figure 29-6A. The electrons in the outer shells are shared by the oxygen (O) atom and the hydrogen (H) atoms so as to make the outer shells of all 3 atoms appear complete. When the outer shells are complete a stable condition exists.

When hydrogen and oxygen combine to form water the force that hold them together is called a COVALENT BOND. An atom of oxygen has a combining power or valence number of 2 because

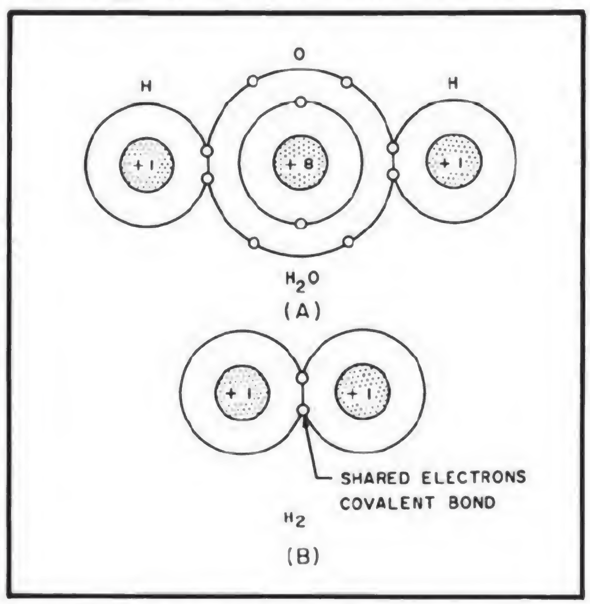

Figure 29-6 - Atoms combining to form molecules.

there are 2 vacancies (2 more electrons needed) in the outer shell. When water is formed, 2 hydrogen atoms share their planetary electrons with the oxygen atom so as to complete the outer shell (a total of 8 electrons) and at the same time the outer shell of each hydrogen atom is completed (2 electrons).

Figure 29-6B illustrates the forming of a hydrogen molecule due to the covalent bonding of two hydrogen atoms.

In certain cases bonds can form between positive and negative ions. Since these bonds are due to electrostatic attraction existing between the two ions this form of bonding is called IONIC BONDING or ELECTROVALENT BONDING. Common table salt is an example of ionic bonding, wherein bonds form between a sodium atom which has given up its single outer electron to become a positive ion and a chlorine atom which has gained an electron to become a negative ion. Due to the fact that semiconductor materials use covalent bonding exclusively, ionic bonding will not be discussed further.

29-9. Pure Germanium Crystal

The carbon atom may be drawn so as to show only those electrons that may be influenced by external forces; that is, those electrons in the outermost shell. The outer shell is represented as an outer concentric circle having 4 electrons spaced equally around it.

In order to fill its outer shell with the maximum number of electrons permitted (8), the carbon atoms in a diamond lattice (a form of crystalline carbon) arrange themselves as shown in Figure 29-7. This arrangement is called a crystal. In the diamond crystal the 4 electrons in the outer shell of one atom are bound as closely to that atom as the 4 electrons in the outer shell of another atom are bound to it. Thus, one atom cannot pull electrons away from another atom; instead, adjacent atoms will share the outer electrons in such a manner that the outer shells are filled to their quota of 8 electrons. Each pair of atoms shares 2 electrons, one from each atom. This pair of electrons (as previously stated for the hydrogen molecule) is called a covalent bond. Thus, to try to fill its outer shell with 8 electrons each carbon atom will establish covalent bonds with 4 other carbon atoms. Each carbon atom shares each of its 4 valence electrons in the outer shell with one other atom, and in return shares an electron in the outer shell of the second atom.

The elements carbon, silicon, and germanium, have the common property of being tetravalent; that is, 4 electrons in the outer shell are able to respond to external forces (enter into chemical reactions). The 4 valence electrons form bonds with 4 electrons from adjacent atoms to form a crystal structure. The atoms arrange themselves in a definite pattern (found in the crystalline forms of carbon, silicon, and germanium). The position of an atom in the crystal is referred to as a lattice site (location).

Germanium and silicon crystallize in the form of a cubic lattice structure. When a pure crystal of germanium or silicon is formed each atom is joined, in the lattice structure, to its four nearest neighbor atoms by covalent bonding. Covalent bonding is accomplished by the four valence electrons of the central atom each coordinating their movement with a valence electron from a neighboring atom. Figure 29-8 illustrates a pure germanium crystal formed by covalent bonding. This coordinated rotation of two valence electrons forms an ELECTRON-PAIR BOND (which is another name for valence bond). These bonds, which hold the atom in place in solid crystalline substances, are stable for certain given numbers of electrons. For example, the bond is especially stable when it contains precisely 2 electrons. The bond is weakened when one electron is removed; it is not strengthened much when a 3rd electron is added.

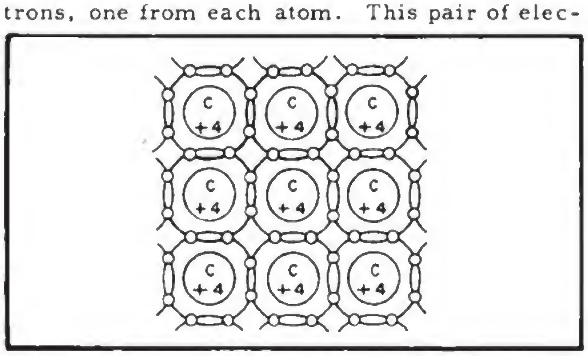

Figure 29-7 - Arrangement of carbon atoms in diamond lattice.

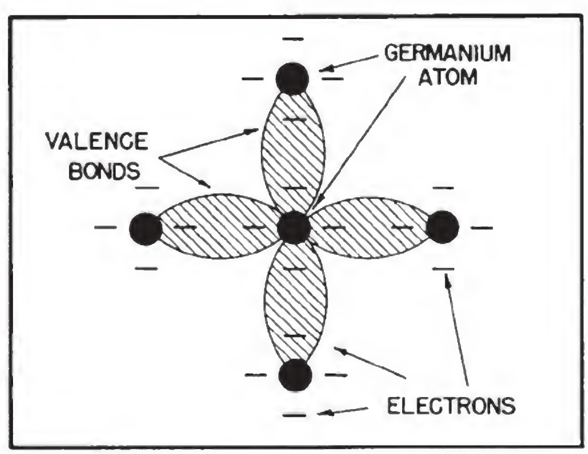

Figure 29-8 - Pure germanium crystal.

A7. Semiconductor.

Q8. If all the valence electrons of a pure germanium or silicon crystal were involved in covalent bonds would the crystal exhibit the properties of a semiconductor or an insulator?

29-10. Intrinsic Conduction

A pure crystal, one containing atoms of only one element such as silicon, is called an INTRINSIC crystal or material.

As mentioned before, a material that is classified as a semiconductor has characteristics between those of a conductor and those of an insulator. The electrons in the outer shell of the atoms of a semiconductor can be removed when some form of energy is applied to the material. The energy may be in the form of heat, light, or an electric field. Then the material will acquire the properties of a conductor. The residual heat contained by a pure germanium or silicon crystal at room temperature is sufficient to cause many of the valence electrons to break their bonds and be elevated to the conduction band. For every electron raised to the conduction band, in an intrinsic material, there will be a vacancy left in the valence band structure. In semiconductor electronics this vacancy is termed a HOLE and for purposes of explanation is said to ACT like a mobile particle with a positive charge equal and opposite to that of an electrons negative charge.

If an intrinsic crystal of germanium or silicon with broken bonds is subjected to a voltage, two kinds of current will flow. The ejected electrons move through the crystal from the negative terminal to the positive terminal, thus constituting ordinary conduction (electron) current. The second current (hole current), however, results from the motion of electrons from one valence bond breaking away to fill up the hole caused by the absence of an electron from an adjacent bond. Thus, holes seem to move in a direction opposite to that of electron flow. Because the hole is a region of net positive charge, the apparent motion is like the flow of particles having a positive charge. An analogy of hole motion is the movement of balls through a tube (Figure 29-9). When ball number one is removed from the tube, a space is left. This space is then filled by ball number 2. Ball number 3 then moves into the space left by ball number 2. This action continues until all the balls have moved one space to the left at which time there is a space left by ball number 8 at the right-hand end of the tube.

The motion of the space is similar to hole motion in the covalent bond structure of semiconductors. The hole motion depends on the MOVEMENT or shifting of VALENCE electrons in the covalent bonds. It is important to remember that holes move ONLY in the VALENCE BAND, also that a hole moves ONLY when a valence band electron moves from one covalent bond to another. The same electrical effect is produced whether electrons move in one direction or holes move in the opposite direction.

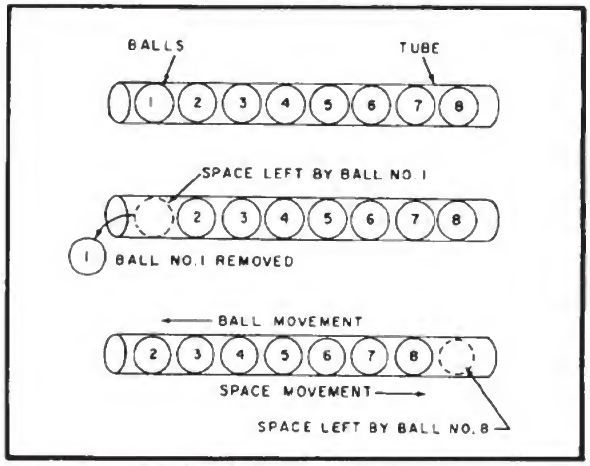

Figure 29-9 - Analogy of hole movement.

The two currents are called ELECTRON CURRENT and HOLE CURRENT. The breaking of a covalent bond that produces a free electron and a hole is called the generation of an ELECTRON-HOLE PAIR. When the charges involved in conduction come from the generation of an electron-hole pair the conduction is INTRINSIC CONDUCTION.

Q9. If five electrons are found in the conduction band of an intrinsic silicon crystal how many holes will be found in the valence band?

Q10. What is a "hole"?

CRYSTAL WITH IMPURITIES

In the pure form, germanium and silicon crystals are of no use as a semiconductor device. However, when a certain amount of impurity is added, the crystal can be made to conduct a current. In order to accomplish this result the quality and quantity of the impurity must be carefully controlled. The added impurities will create either an excess or a deficiency of electrons depending on the kind of impurity added.

The impurities that are important in semiconductor materials are those impurities that align themselves in the regular lattice structure

whether they have one valence electron too many, or one valence electron too few. The first type loses its extra electron easily and in so doing increases the conductivity of the material by contributing a free electron. This type of impurity has 5 valence electrons and is called a PENTAVALENT impurity. Arsenic, antimony, bismuth, and phosphorous are pentavalent impurities. Because these materials give up or donate one electron to the material they are called DONOR impurities.

The second type of impurity tends to compensate for its deficiency of 1 valence electron by acquiring an electron from its neighbor. Impurities of this type in the lattice structure have only 3 valence electrons and are called TRIVALENT impurities. Aluminum, indium, gallium, and boron are trivalent impurities. Because these materials accept one electron from the material they are called ACCEPTOR impurities.

Semiconductors that have no impurities are called intrinsic semiconductors. Semiconductors that have either acceptor or donor impurities are called extrinsic semiconductors.

The energy levels of the donor and acceptor impurities are slightly different than those of the intrinsic (germanium or silicon) material. At a temperature of absolute zero, which is zero degrees Kelvin, the energy level of the excess donor electron exists just BELOW the conduction band. At zero degrees Kelvin the energy level of the acceptor impurities is just ABOVE the valence band. Figure 29-10 illustrates the donor and acceptor impurity energy levels in relation to the intrinsic energy levels, by use of an energy level diagram.

NOTE: The forbidden energy gap applies ONLY to electrons associated with the germanium or silicon atoms and not to the electrons associated with the impurity atoms.

29-11. N-Type Germanium

When a pentavalent impurity like arsenic is added to germanium it will form covalent bonds with the germanium atoms. Figure 29-11 illustrates an arsenic atom (As) in a germanium lattice structure. The arsenic atom has 5 valence electrons in its outer shell but uses only 4 of them to form covalent bonds with the germanium atoms, leaving one electron relatively free in the crystal structure. Because this type of material conducts by electron movement it is called a negative-carrier type or N-type semiconductor. Pure germanium may be converted into an N-type semiconductor by "doping" it with any element containing 5 electrons in its outer shell. Other pentavalent elements which may be used in place of arsenic as "dopants" are phosphorous, antimony, and bismuth. The amount of the impurity added is very small; it is of the order of one atom of impurity in 10 million atoms of germanium.

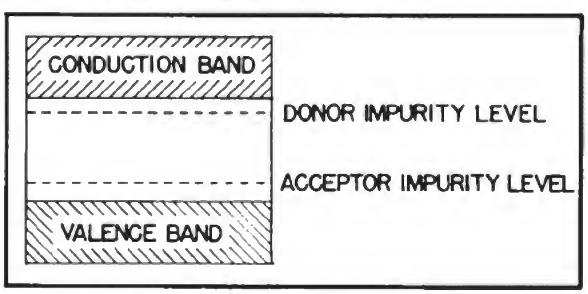

Figure 29-10 - Energy levels for donor and acceptor impurities.

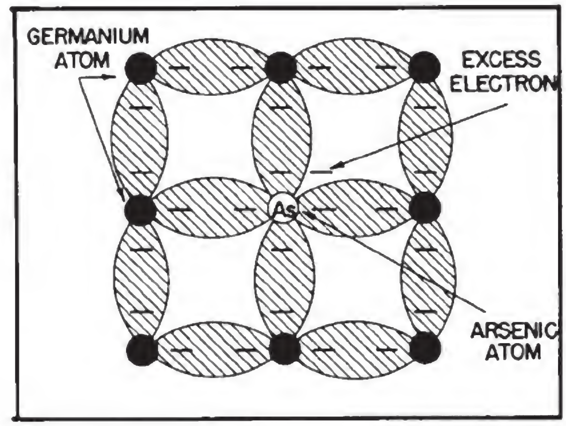

Figure 29-11 - Donor impurity.

Q11. Does the freeing of the excess donor electron require more or less energy than the freeing of an intrinsic valence electron (in an N-type material)?

NOTE: Intrinsic refers to the original or pure material of the crystal (germanium or silicon).

29-12. P-Type Germanium

A trivalent impurity element can also be added to pure germanium to "dope" the material. In this case the impurity has one less electron than it needs to establish covalent bonds with 4 neighboring atoms. Thus in one covalent bond there will be only one electron instead of 2. This arrangement leaves a vacancy in that covalent bond.

Figure 29-12 shows the germanium lattice structure with the addition of an indium atom (symbol In). The indium atom has one electron less than it needs to form covalent bonds with the 4 neighboring atoms. Gallium and boron also exhibit these characteristics.

A8. Insulator. If all the valence electrons are in bonds there will be no electrons in the conduction band (this situation could only exist at absolute zero temperature).

A9. Five holes.

A10. A hole is a vacancy in the valence band structure of a semiconductor material.

A11. Much less. The intrinsic valence electron must jump the entire forbidden energy gap whereas the energy level of the excess donor electron exists just below the conduction band and requires very little energy to become a free electron.

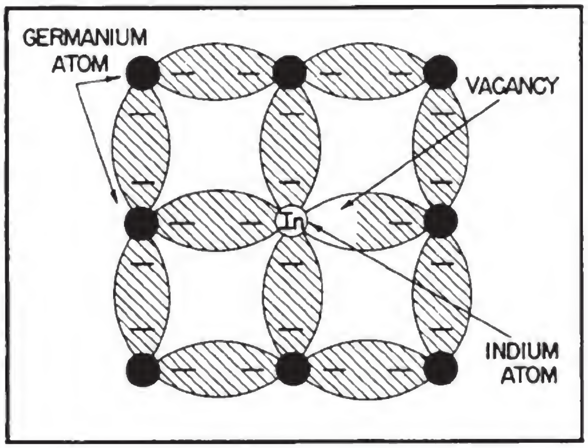

Figure 29-12 - Acceptor impurity.

It must be noted that the acceptor atom is neutral (the number of electrons in orbit equal the number of protons in the nucleus) when inserted into the crystal. Therefore, the acceptor cannot attract an electron from the germanium or silicon atoms. However, should thermal energy unleash an electron from a neighboring silicon or germanium atom and bring the electron favorably close to the acceptor atom, the acceptor atom will capture or "accept" the electron. It must be stressed that this capture occurs NOT as a result of electrostatic attraction (the acceptor atom is neutral), but, as a result of a natural tendency of atoms to form complete covalent bonds. The energy levels of the acceptor atoms exist just above the valence band of the crystal. While the valence electron of an excited silicon or germanium atom may not possess enough energy to be elevated to the conduction band it will possess enough energy to be elevated to the energy level of the acceptor atom. It may be said that the acceptor atom "stores" an electron from the valence band structure of the crystal. The instant a germanium or silicon valence electron leaves the bonding structure for "storage" a hole is "created" in the valence band. Thus, a hole results in a positively ionized germanium or silicon atom.

Since in a P-type material an electron does not have to be raised to the conduction band in order to create a hole THERE WILL BE MANY MORE HOLES IN THE VALENCE BAND THAN THERE ARE ELECTRONS IN THE CONDUCTION BAND. Because this semiconductor material conducts by the apparent movement of holes which are considered to be positive charges, it is called a positive-carrier type or P-type semiconductor.

When a valence electron moves to fill a hole the hole APPEARS to move to the spot previously occupied by the valence electron. Figure 29-13 illustrates the movement of a hole through a crystal.

In Figure 29-13 the original position of the hole is at point one (P_1). One of the valence electrons from point two (P_2) for instance, moves over and fills the hole. In moving to fill the hole the electron leaves a hole at P_2. Thus, it appears as if the hole has moved from P_1 to P_2. The above action is repeated and the hole moves through the valence structure of the crystal in a random manner until it finally arrives at P_3. Therefore, the movement of a hole through the crystal is dependent on the movement of valence band electrons.

29-13. Charges in N and P-Type Materials

When a donor material such as arsenic is added to germanium, the fifth electron in the outer ring of the arsenic atom does not become a part of a covalent bond. This extra electron (when acted on by some force) may move away

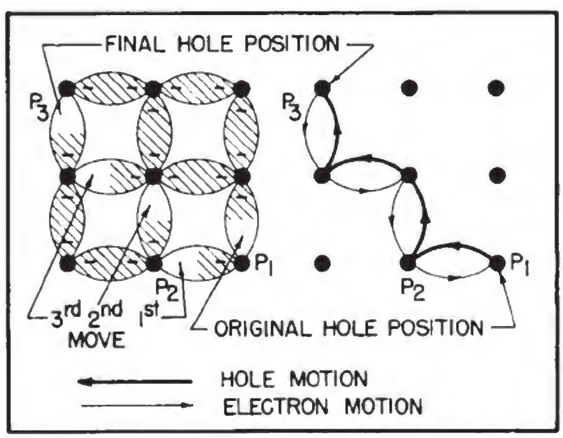

Figure 29-13 - Hole movement through crystal.

from the arsenic atom to one of the nearby germanium atoms in the N-type material.

NOTE: In the discussion to follow the electrons of the inner shells, and the protons needed to neutralize them, will be neglected. Only the electrons of the outermost shell and the corresponding protons will be referred to.

The arsenic atom has a positive charge of 5 units (5 protons) in the nucleus and when the excess electron moves away from the arsenic atom there will be only 4 electrons left to neutralize the positive charges. As a result there will be a region of positive charge around the arsenic atom.

Although there is a region of positive charge around the arsenic atom after the electron has moved away, and a region of negative charge around the freed excess electron, the total charge on the N-type crystal remains the same. In other words the total charge is zero. There are exactly enough electrons to neutralize the positive charges on the nuclei of all the atoms in the crystal. However, because some of the electrons may move about in the crystal, there will be regions in the crystal where there are negative charges and other regions where there will be positive charges, even though the net charge on the crystal is zero.

In a P-type material having an impurity such as indium added to it, a similar situation may exist. Indium has only 3 electrons in its outer ring. Three electrons are all that are needed to neutralize the positive charges of the 3 protons in the nucleus. However, with only 3 electrons in the outer shell, there is a vacancy in one of the covalent bonds formed between the indium atom and the 4 adjacent germanium atoms. If an electron moves in to fill this vacancy there will be one more electron in the indium atom than is needed to neutralize the positive charge of 3 units. Thus there will be a region of negative charge around the indium atom.

Similarly, when the germanium atom gives up an electron to fill in the vacancy in the indium covalent bond, the germanium atom will be short an electron and there will be a region of positive charge around this germanium atom. While the giving up of an electron by a germanium atom and the acquisition of an electron by the indium atom charges (ionizes) both atoms involved, the net charge on the P-type crystal is still zero. There is simply one atom that is short an electron and another atom that has one too many. The crystal itself does not acquire any charge.

These ionized atoms produced in both N and P-type germanium are not concentrated in any one part of the crystal, but instead are spread uniformly throughout the crystal. If any region within the crystal were to have a very large number of positively charged atoms, these atoms would attract free electrons from other parts of the crystal to neutralize part of the charged atoms so that the charge would spread uniformly throughout the crystal. Similarly, if a large number of atoms within a small region had an excess of electrons, these electrons would repel each other and spread throughout the crystal.

Both holes and electrons are involved in conduction. The holes are called positive carriers and the electrons negative carriers. The one present in the greatest quantity is called the MAJORITY CARRIER; the other is called the MINORITY CARRIER. In N-type material the electrons are the majority carriers and holes the minority carriers. In P-type material the holes are the majority carriers and electrons the minority carriers.

Q12. Do majority and minority carriers both exist in an N-type material?

29-14. Current Flow in N-Type Material

Current flow through an N-type material is illustrated in Figure 29-14. Conduction in this type of semiconductor is similar to conduction in a copper conductor. That is, the application of voltage across the material will cause the donor impurity electrons in the conduction band to migrate through the crystal toward the positive potential point.

However, certain differences exist between the N-type semiconductor and a copper conductor. For example, the semiconductor resistance decreases with temperature increase, because more carriers are made available at higher temperatures. Increasing the temperature will generate more electron-hole pairs (by breaking germanium or silicon covalent bonds), elevating many electrons to the conduction band and thus causing increased conductivity (decreased resistance). In the copper conductor, increasing the temperature does not increase

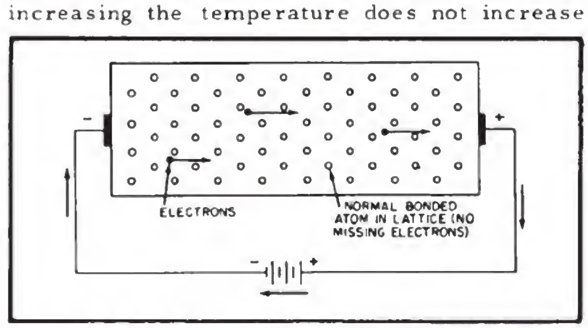

Figure 29-14 - Current flow in N-type material.

A12. Yes. Although the majority carriers, which are electrons in the N-type material, far outnumber the minority carriers, some holes do exist due to the generation of electron hole pairs.

---

the number of carriers but increases the thermal agitation or vibration of the structure so as to impede the current flow further (increase the resistance).

29-15. Current Flow in P-Type Material

Current flow through a P-type material is illustrated in Figure 29-15. Conduction in this material is by positive carriers (holes). In order that the hole appears to move, an electron in a nearby lattice site must shift to the position where the hole existed originally. Thus the hole moves from the positive terminal to the negative terminal of the P-type material. Electrons from the external circuit enter the negative terminal and cancel holes in the vicinity of the terminal while at the positive terminal, electrons are being removed from the covalent bonds, thus creating new holes. The new holes then move toward the negative terminal (the electrons shifting to the positive terminal) and are cancelled by more electrons emitted from the negative terminal. This process continues as a steady stream of holes (hole current) moving toward the negative terminal.

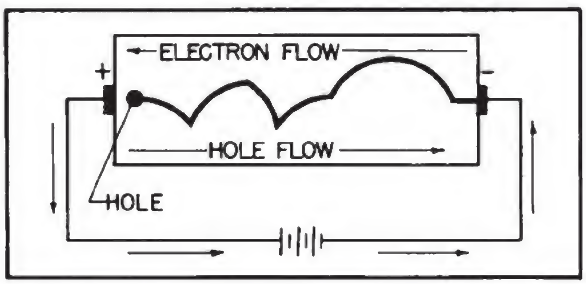

Figure 29-15 - Current flow in P-type material.

In both N-type and P-type materials, current flow in the EXTERNAL CIRCUIT consists of electrons and is out of the negative terminal of the battery and into the positive terminal.

Q13. With a battery connected across a piece of N-type material would the current flow equally well if the polarity of the battery were reversed?

Q14. What would be the effect of a large increase in temperature on P-type material?

29-16. PN Junction

A semiconductor diode is made by taking a single crystal (for example, germanium) and adding a donor impurity to one region and an acceptor impurity to the other. This gives a single crystal with an N section and a P section. Where the two sections meet is a junction. Contacts are fastened to the two ends of the crystal. The result is a simple PN junction or junction diode.

One portion of the crystal is P-type material; this is the portion containing the acceptor impurity. The other portion is N-type material. This region contains the donor impurity. The end contacts are large surfaces that make a good connection with the crystal. If the connections were not good there might be rectifying properties where they come in contact with the crystal.

Figure 29-16 is a pictorial representation of a PN junction.

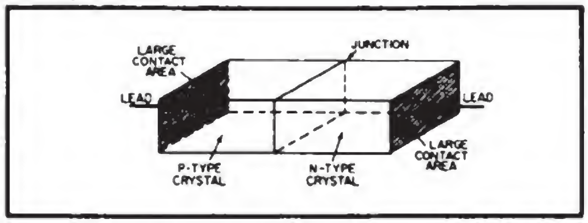

Figure 29-16 - PN junction pictorial diagram.

An isolated piece of N-type material is electrically neutral; that is, for every free electron in the conduction band there is a positively ionized donor atom in the crystal lattice structure. Thus, while there is an abundance of free negative charges each one is balanced by a fixed positive charge and the overall charge of the crystal is zero.

An isolated piece of P-type material is also electrically neutral; for every hole (positively ionized germanium or silicon atom) there exists a negatively ionized acceptor atom. Thus, the overall charge of the P-type crystal is zero. Figure 29-17 shows an electrical representation of an isolated N and P-type material with balanced charges.

Figure 29-17 is used to represent the even distribution of carriers and ions throughout the crystals. The carriers are placed beside each ion to indicate the balancing of positive and negative charges.

While there are many and varied methods of combining a P and N-type material into a PN junction, there is one qualification that must be met. The junction, when completed, must have the properties of a single crystal.

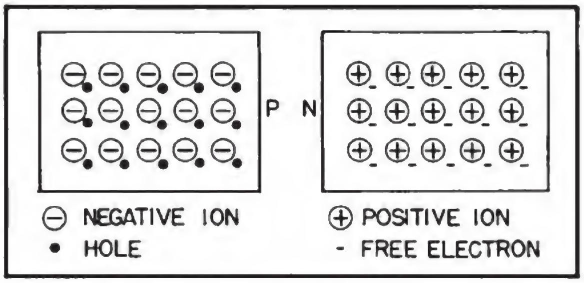

Figure 29-17 - Isolated P and N-type materials.

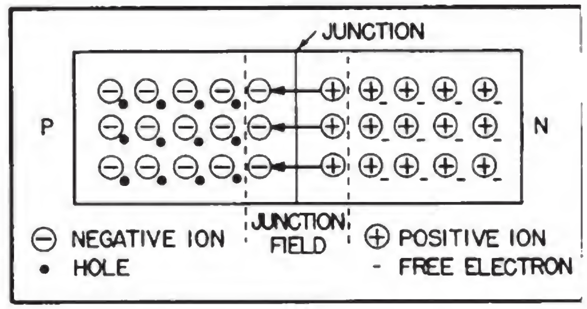

Figure 29-18 - PN junction field.

29-17. Junction Field

Regardless of the manner in which it is accomplished, once the junction is formed an interesting action takes place.

The electrons in the conduction band of the N-type material "see" the relatively empty conduction band of the P-type material and begin to DIFFUSE (move or spread out) across the junction into the P material. The free electrons lose energy, fall into the valence band and recombine with the holes. This recombination eliminates a free electron and a hole. When the electron moved from the N material and diffused across the junction it left a positive ion which is no longer balanced by a negative charge. This same electron, upon recombination with a positive germanium ion, eliminates a hole in the P material. The elimination of this hole and its associated positive charge leaves a negative acceptor ion in the P material which is no longer balanced by a positive charge.

These ions are fixed in place in the crystal lattice structure and cannot move. Thus, they make up a layer of fixed charges on the two sides of the junction. On the N side of the junction there is a layer of positively charged ions; on the P side of the junction there is a layer of negatively charged ions. An electrostatic field is established across the junction between the oppositely charged ions. Figure 29-18 illustrates the electrostatic field of the junction, called the JUNCTION FIELD. The junction field in Figure 29-18 is greatly exaggerated for purposes of explanation. The diffusion of electrons across the junction will continue until the magnitude of the electrostatic field is increased to the point where the electrons no longer have enough energy to overcome it. At this point an equilibrium condition is established and for all practical purposes the migration of carriers across the junction ceases. The action just described occurs almost instantly when the junction is formed. Only the carriers in the immediate vicinity of the junction are involved in the forming of the junction field. The carriers throughout the remainder of the N and P materials are relatively undisturbed and maintain the balanced condition of positive and negative charges in their respective materials.

Figure 29-18 also indicates that there are no mobile carriers within the junction field. Any mobile carriers which would happen to appear within the field would be accelerated out to the boundaries of the field. Electrons in an electrostatic field are accelerated AGAINST the arrow and holes are accelerated WITH the arrow. Since the junction field has a depletion of mobile carriers it is often called the DEPLETION LAYER or DEPLETION REGION. Other names for the junction field are the SPACE CHARGE REGION and the JUNCTION BARRIER.

The charge on the impurity atoms is distributed across the PN junction as shown in Figure 29-19, curve 1. In the P region the ionized acceptors have a negative charge, and in the N region the ionized donor atoms have a positive charge. At the junction the charge is zero.

However, in the P region there are holes which have a positive charge and in the N region there are free electrons which have a negative charge. The distribution of holes and free electrons is shown by curve 2 (Figure 29-19).

The potentials at the junction have driven the holes away from the junction in the P region and the electrons away from the junction in the N region so that the charges in the P region and in the N region are moved further apart. Thus the slope of curve 2 is more gradual than that of curve 1. The charge at the junction is zero but the rise on either side is more gradual than that of curve 1. Moving further into the P region (curve 2) the charge becomes positive due to the holes; moving into the N side of the crystal the charge is negative due to the electrons.

The net charges on the crystal in the P region is equal to the difference between the charge on the ionized acceptor atoms and the charges

A13. Yes. N-type material alone has no rectifying qualities.

A14. The resistance would decrease due to the addition of many carriers of hole current. (An increase of temperature will generate many electron-hole pairs).

on the holes. The net charge on the crystal in the N region is equal to the difference between the charge of the ionized donor atoms and the electrons. These charges cancel except in the immediate region of the junction. This arrangement is indicated by curve 3 which is the sum of curves 1 and 2 (Figure 29-19).

Since the areas of the P and N materials beyond the junction field have zero potential, due to the equality of oppositely charged mobile carriers and ions, only the field of the PN junction is shown in Figure 29-19.

In the area near the junction there is a negative charge in the P region and a positive charge in the N region. As stated earlier, they act as a barrier to prevent further diffusion of holes from the P region into the N region and the diffusion of electrons from the N region into the P region. This potential barrier is a potential difference (or voltage) across the junction and is of the order of a few tenths of a volt. The physical distance from one side of the field or barrier to the other is referred to as the BARRIER WIDTH. The width of the barrier, with

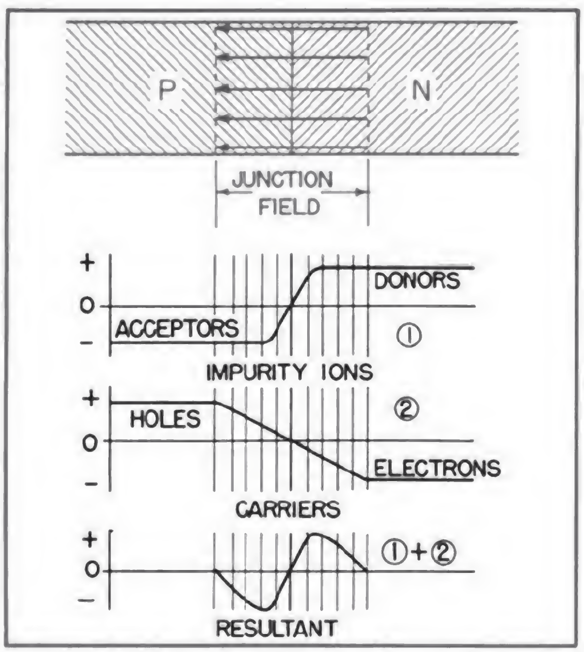

Figure 29-19 - Junction field potentials.

no external potential applied, depends on the density of mobile carriers in the crystal. The BARRIER HEIGHT or POTENTIAL HILL is the difference of potential across the depletion region (intensity of the junction field).

Q15. If a voltmeter (with a very high input impedance) was placed across a piece of plain N-type material what would be the indication on the meter?

Q16. Why does the N-type material have positive ions?

29-18. Intrinsic Conduction in a PN Junction

A limitation imposed on the junction diode is the result of electron-hole pairs that are being formed at random within the crystal due to energy imparted to the crystal by heat, light and electromagnetic radiation. Away from the depletion layer these carrier pairs will recombine without materially affecting the carrier concentration in the crystal. In other words the holes will remain the majority carriers in the P material and electrons will remain the majority carriers in the N material.

However, as previously mentioned, both holes and electrons are involved in conduction at all times. There are minority carriers in both regions; holes in the N material and electrons in the P material. The holes produced in the N material near the junction are minority carriers. The direction of the junction field is such that minority carriers are readily accelerated and pass across the junction. These holes will tend to neutralize the negative ions on the P side of the junction. Similarly free electrons produced on the P side of the junction will pass across the junction and neutralize positive ions on the N side of the junction. This action is an example of intrinsic conduction which is undesirable. This is called intrinsic conduction because the electrons and holes involved come from the generation of electron-hole pairs and not from the impurity atoms.

BIASED PN JUNCTIONS

If a battery is connected across the PN junction the battery potential will bias the junction. If the battery is connected so that its voltage opposes the barrier potential across the junction it will reduce the intensity and width of the junction field and thereby aid current flow through the junction and the junction is said to be biased in the FORWARD direction (low resistance). If the battery is connected across the junction so that its voltage aids the barrier potential across the junction it will increase the

intensity and width of the junction field and thereby oppose current flow through the junction and the junction is said to be REVERSE-BIASED or biased in the reverse direction. This is the direction of high resistance.

29-19. Forward Bias

The forward bias connection is illustrated in Figure 29-20. Here the positive terminal of the bias battery is connected to the negative side of the barrier potential (P-type side of the junction) and the negative terminal of the battery is connected to the positive side of the barrier potential (N-type side of the junction).

The positive terminal of the battery connected to the end of the P-type germanium, repels holes toward the junction. The holes are repelled toward the junction because germanium valence electrons are being attracted by the positive terminal of the battery. The holes moving toward the junction neutralize the negatively charged acceptor ions at the edge of the junction field.

On the N side of the crystal the negative terminal of the battery repels electrons toward the junction. These electrons neutralize the positively charged donor atoms at the edge of the junction field. Since ions on both sides of the field are being neutralized the field will be decreasing in both width and intensity. Thus, the effect of the battery bias voltage in the forward direction is to reduce the barrier potential across the junction and to allow majority carriers to cross the junction.

The current flow and method of conduction in a forward biased PN junction is as follows:

An electron leaves the negative terminal of the battery and moves to the terminal of the N-type material. It enters the N material, where it is the majority carrier, and moves to the edge of the junction field. Due to forward bias the field now offers less opposition to the electrons and they pass through the depletion region into the P-type material. The electron loses energy in overcoming the opposition of the junction field and upon entering the P material encounters a concentration of holes at the edge of the field. The electron falls in energy and combines with one of the holes. By virtue of hole conduction the electron (now a valence electron) moves through the P material toward the terminal. Upon reaching the terminal it is attracted out of the P material by the positive terminal of the battery and proceeds through the connecting wire to the positive battery terminal and the completion of its journey.

It is important to remember that in the forward biased junction condition, conduction is by the majority carriers (holes in the P-type material and electrons in the N-type material). Increasing the battery voltage will increase the number of majority carriers arriving at the junction and the current flow increases. The only limit to current flow is the resistance of the material on the two sides of the junction. If the battery voltage is increased to the point where the barrier potential across the junction is completely neutralized heavy current will flow and the junction may be damaged from the resulting heat. Therefore, the voltage of the bias battery is limited to a relatively small voltage.

29-20. Reverse Bias

With reverse bias applied to the junction diode the negative terminal of the battery is connected to the P-type material and the positive terminal connected to the N-type material. The negative terminal of the battery attracts the holes away from the edge of the junction field on the P side while the positive terminal attracts the electrons away from the edge of the field on the N side. This action increases the barrier potential across the junction because there are fewer holes on the P side of the junction to neutralize the negative ions, and fewer electrons on the N side to neutralize the positive ions formed on this side of the junction. The increase in barrier potential prevents current flow across the junction by MAJORITY carriers.

Figure 29-21 illustrates a reversed biased PN junction.

Notice that the width of the junction field has been increased. The concentration of the mobile carriers at the terminals has left many

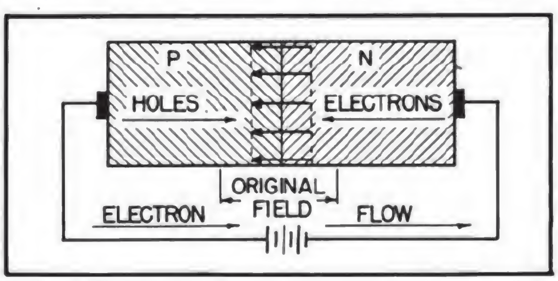

Figure 29-20 - Forward biased PN junction.

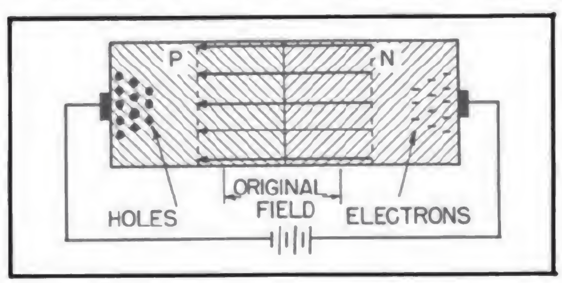

Figure 29-21 - Reversed biased PN junction.

774-639 O - 65 - 2

A15. Zero. Plain N or P type material is electrically neutral.

A16. In forming the N type material the donor atom gives up one of its valence electrons, leaving a donor ion with one more proton than electron.

---

ions unneutralized at the edge of the field, therefore, the field will extend to include these ions.

The current flow across the barrier is not zero, however, because of minority carriers crossing the junction. Holes forming in the N material are repelled toward the junction and electrons being released in the P material are repelled (by the battery) toward the junction. In both cases these are MINORITY carriers. The junction field is in opposition to MAJORITY carriers only. Therefore, when the minority carriers reach the junction field they will easily be accelerated across the junction. This situation was described as intrinsic conduction due to electron-hole pairs continually forming at random within the crystal (before any bias is applied) as a result of the energy of the crystal.

Thus, under reverse bias conditions there will be a small current flow due to minority carriers crossing the junction. This current flow is small at normal operating voltages and temperatures.

Q17. Is the potential hill of a PN junction increased or decreased when the junction is forward biased?

Q18. If a PN junction is biased in such a manner as to increase the depletion region, will the junction field aid or oppose minority current flow?

29-21. PN Junction Symbol and Ratings

The schematic symbol of a PN junction diode is shown in Figure 29-22. The bar represents the cathode and the arrow represents the anode. Electron flow is against the arrow. For clarification a pictorial diagram of a PN junction is also illustrated.

PN junction diodes are generally rated for the MAXIMUM AVERAGE FORWARD CURRENT, the PEAK RECURRENT FORWARD CURRENT, the MAXIMUM SURGE CURRENT, and the PEAK REVERSE VOLTAGE (PRV).

MAXIMUM AVERAGE FORWARD CURRENT is given at a specified ambient temperature and

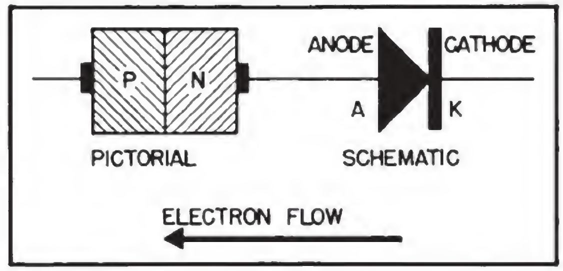

Figure 29-22 - PN junction symbols.

refers to the maximum amount of average current which can be permitted to flow in the forward direction. Typical values of current for silicon diodes range from 150 ma to as high as 250 amperes for a single diode.

PEAK RECURRENT FORWARD CURRENT is the maximum peak current which can be permitted to flow in the forward direction in the form of recurring pulses.

MAXIMUM SURGE CURRENT is the maximum current permitted to flow in the forward direction in the form of non-recurring pulses. Current should not equal this value for more than a single forward cycle of diode operation.

PEAK REVERSE VOLTAGE is one of the most important ratings. PRV indicates the maximum reverse bias voltage which may be applied to a diode. All of the above ratings are subject to change with temperature variations. If the temperature increase is above that stated for the ratings, all the ratings must be degraded by an appropriate amount.

Reverse bias (also called back bias) applied across a junction diode increases the barrier potential, making it more difficult for majority carriers to cross the junction. However, some minority carriers will still cross the junction, with the result that there will be a small current. This action is indicated in the static voltage-current characteristic curve, shown in Figure 29-23. The forward portion of the curve indicates that the diode conducts easily when the potential across the junction is in the direction of forward bias (P side positive and N side negative). The diode conducts poorly in the high resistance direction (reverse bias, P side negative and N side positive). For this condition the holes and electrons are drawn away from the junction, causing an increase in the barrier potential. However, if the reverse bias is increased beyond a critical value the reverse current increases rapidly due to avalanche breakdown. The peak reverse voltage rating

ranges from a low of approximately 40 volts to a high of 1000 volts for general applications, with special types (CR212, etc.) extending to approximately 12,000 volts.

Avalanche breakdown occurs when the applied voltage is sufficiently large to cause the covalent bond structure to break down. At this point a sharp rise in reverse current occurs. The acceleration of the few holes and electrons continues to such a point that they have violent collisions with the valence bond electrons of the germanium or silicon crystal atoms releasing more and more carriers. The maximum reverse voltage of the semiconductor diode corresponds to the peak inverse voltage of an electron-tube diode.

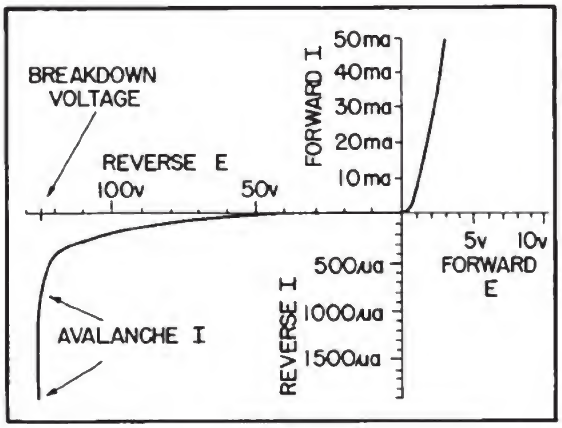

Figure 29-23 - Voltage-current characteristics of germanium diode IN63(25°C).

Q19. Is the cathode of a PN junction made of N or P type material?

Q20. The normal maximum average forward current rating of a silicon diode is 500 ma at 25°C. If the diode is operated in an ambient temperature of 85°C would the diode current rating increase or decrease?

PN JUNCTION RECTIFIERS

29-22. PN Junction Half-Wave Rectifier

It has been shown that a PN junction presents a low resistance to current flow in one direction and a high resistance to current flow in the opposite direction. Thus, the PN junction or semiconductor diode possesses the property of rectification.

Many types of semiconductor diodes are available. They vary in size from tiny ones hardly bigger than a pinhead, used in subminiature circuitry, to large 500 ampere rectifiers used in power supplies. A few of the many types are pictured in Figure 29-24.

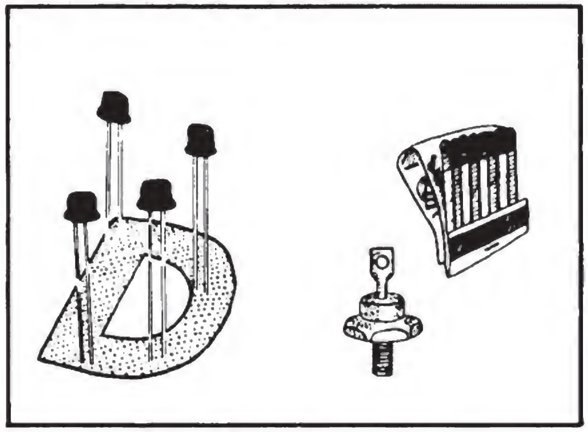

Figure 29-24 - Semiconductor diodes.

The semiconductor diode may be used to replace the electron tube diode in the familiar power supply configurations.

A simple half-wave rectifier circuit utilizing a silicon diode (PN junction) is illustrated in Figure 29-25. In operation an alternating voltage is applied to the primary of transformer (T_1). The transformer windings are wound so that there is no phase reversal. Thus, when the top of the primary is made positive by the input signal the induced voltage will make the top of the secondary positive. Forward biasing of a PN junction occurs when a positive voltage is applied to the P-type material and a negative voltage is applied to the N-type material. When the input signal causes the top of the secondary winding to become positive the diode is forward biased and will conduct. Electrons leave the bottom or negative end of the winding, flow through R_L, through the PN junction and back to the top of the transformer winding. This current flow causes a voltage drop across the load resistor with a polarity negative at the bottom and positive at the top. The input and output waveforms (Figure 29-25) show a positive voltage being produced for the first half cycle of the input signal.

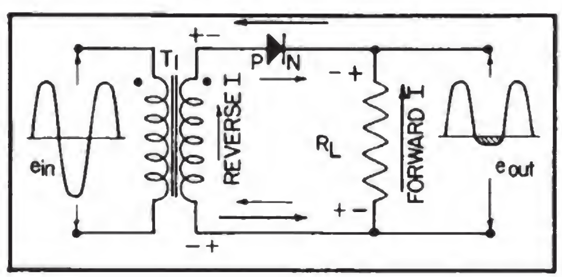

Figure 29-25 - PN junction half-wave rectifier.

A17. Decreased. Forward bias reduces the width and intensity of the junction field.

A18. Aid. The direction of the junction field is such that it will ALWAYS aid minority current flow. The polarity and magnitude of bias merely determines the DEGREE of aid.

A19. N-type.

A20. Decrease. The maximum ratings will be much less with increased temperature.

For the negative half cycle of the input signal the top of the secondary winding will become negative. Applying the negative potential to the P material of the junction will reverse bias the diode and no majority current will flow. However, a small amount of reverse current will flow in the circuit due to the minority current flow in the PN junction. This is indicated in the output waveform by the small negative voltage below the zero line. Under normal operating conditions the reverse current is a VERY small percentage of the forward current.

NOTE: Reverse current in the illustrations is overemphasized for explanatory purposes.

29-23. PN Junction Full-Wave Rectifier

A full-wave rectifier permits current to flow in the same direction through the load during both alternations of an ac input signal.

Two silicon diodes may be used in a full-wave rectifier configuration to obtain full-wave rectification. Such a configuration is illustrated in Figure 29-26.

During operation, an alternating signal is applied to the primary of transformer (T_1). When the input signal is going positive it makes the top of T_1 secondary positive and the bottom negative. Diode two (CR2) is forward biased because its anode (P material) has the positive potential from the top of the secondary applied to it. Diode one (CR1) is reversed biased because its anode has the negative potential from the bottom of the secondary applied to it. Thus, when the input signal is going positive CR2 will be conducting and CR1 will not. The current path will be into ground at the secondary winding center tap, up through the load resistor (R_L), through CR2 and back to the top of the winding. The waveforms (Figure 29-26) for the current through the diodes show CR2 conducting by MAJORITY current flow and CR1 conducting by MINORITY current flow during the positive

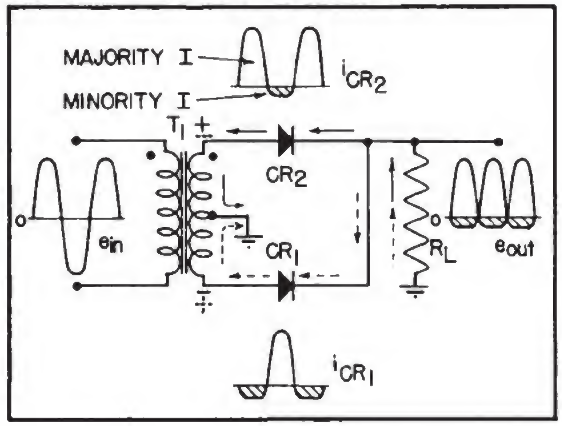

Figure 29-26 - PN junction full-wave rectifier.

half of the input signal. Majority current flows upward through R_L while minority current flows downward. Therefore, the minority current will subtract a small amount from the majority current. This is indicated in the output waveform. The illustration is greatly exaggerated for explanation purposes. Under normal operating conditions the reverse current is so small that it subtracts a negligible amount from the forward current.

On the negative half cycle of the input signal CR1 will be forward biased and CR2 will be reversed biased. The forward current path will be into ground at the center tap of the secondary winding, up through the load resistor, through CR1 and back to the bottom of the winding. Reverse current will flow from the top of the winding through CR2, down through the load resistor, and through ground to the center tap of the winding. Notice that the output waveform has the same polarity regardless of which diode is conducting. The output of the circuit will be a pulsating dc.

Q21. What polarity output voltage would the full-wave rectifier have if both silicon diodes were reversed?

29-24. PN Junction Bridge Rectifier

If four silicon diode rectifiers are connected as shown in Figure 29-27, the circuit is called a BRIDGE RECTIFIER. The input waveform (Figure 29-27), is applied to diagonally opposite corners of the network, points X and Y. The output is taken from the remaining two corners, across the load resistor R_L. During the first half cycle of the applied alternating voltage, point X becomes negative with respect to point Y by the amount of voltage induced in the secondary of the transformer. During this time,

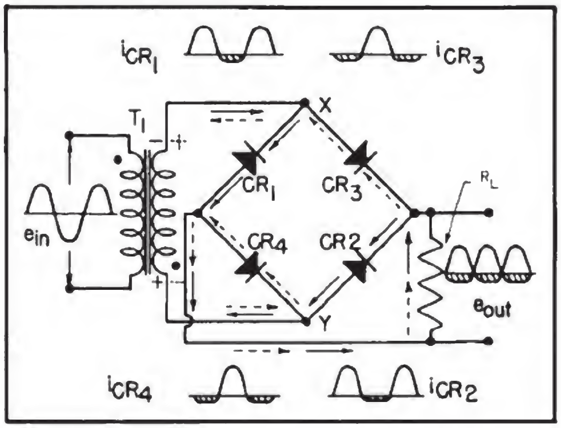

Figure 29-27 - PN junction bridge rectifier.

the voltage between points X and Y may be considered to be impressed across a load consisting of CR1, load resistor R_L, and CR2 in series. The voltage applied across these diodes makes their anodes more positive than their cathodes (forward bias) and forward current flows in the path indicated by the solid arrow. The waveforms (Figure 29-27) indicate the voltage and current conditions in the circuit for the negative half cycle of the input signal. The current through the load resistance is always in the same direction.

On the second or negative half cycle, point X becomes positive with respect to point Y. During this time the voltage between points X and Y may be considered to be impressed across a load consisting of CR4, R_L, and CR3 in series. Since the forward voltage drop of the silicon diodes is very small most of the voltage will appear across R_L. The current path is indicated by the dotted arrows.

The bridge rectifier is a full-wave rectifier because current flows in the loading during both alternations of the input signal. One advantage of the bridge rectifier over the conventional full-wave rectifier is that, with a given transformer, the bridge circuit produces a voltage output nearly twice that of the full-wave circuit. This increase in voltage is due to the fact that the bridge circuit does not use a center tapped secondary as does the full-wave circuit. Therefore the bridge circuit applies the full voltage of the secondary to the rectifiers while the full-wave circuit applies only the voltage between the center tap and one end of the secondary winding.

A second advantage of the bridge circuit is that the peak reverse voltage across a diode is only half the peak reverse voltage across a diode in a full-wave circuit designed for the same output voltage.

Q22. In a full-wave rectifier circuit, what is the effect on the junction field of the non-conducting diode when the other diode is conducting?

Q23. What is the cause of the reverse current flow in the PN junction bridge rectifier.

FILTER CIRCUITS

The preceding paragraphs have discussed methods of converting alternating current into pulsating direct current by use of various solid state rectifier circuits. Most electronic equipment requires a smooth dc supply, approaching the ripple-free output of a battery. Conversion of pulsating direct current to pure direct current is accomplished by the use of properly designed filters.

The unfiltered output of a full-wave rectifier is shown in Figure 29-28. The polarity of the output voltage does not reverse, but its magnitude fluctuates about an average value as the successive pulses of energy are delivered to the load. In Figure 29-28, the average voltage is shown as the line that divides the waveform

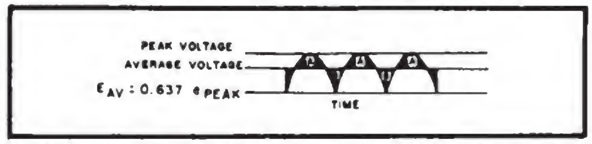

Figure 29-28 - Unfiltered output voltage of a full-wave rectifier.

so that area A equals area B. The fluctuation of voltage above and below this average value is called RIPPLE. The frequency of the main component of the ripple for a full-wave rectifier is twice the frequency of the voltage that is being rectified. In the case of the half-wave rectifier the ripple has the same frequency as the input alternating voltage. Thus, if the input voltage is obtained from a 60 cps source, the main component of ripple in the output of a half-wave rectifier is 60 cps, and in the full-wave rectifier it is 120 cps.

The output of any rectifier is composed of a direct voltage and an alternating or ripple voltage. For most applications, the ripple voltage must be reduced to a very low amplitude. The amount of ripple that can be tolerated varies with different applications.

The PERCENTAGE OF RIPPLE is 100 times the ratio of the RMS value of the ripple voltage at the output of a rectifier filter to the average value, E_o, of the total output voltage. Figure 29-29 indicates graphically how the percentage of ripple may be determined. It is assumed that the ripple voltage is sinusoidal.

A21. Negative. Forward current flow would be downward through the load resistor.

A22. The junction of the non-conducting diode is increased because it is reverse biased at the time the other diode is conducting.

A23. The generation of electron-hole pairs in the silicon diodes.

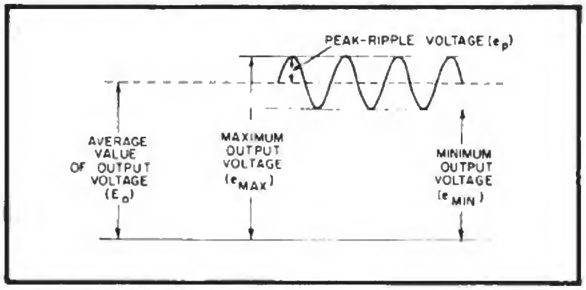

Figure 29-29 - Percentage of ripple.

The formula for determining the percentage of ripple is:

$$\text{percentage of ripple} = \frac{E_{RMS}}{E_o} \times 100 \qquad (17\text{-}1)$$

where $E_{RMS} = 0.707$ of e_p, and e_p is the peak value of the ripple voltage.

A circuit that eliminates the ripple voltage from the rectifier output is called a FILTER. Filter systems in general are composed of a combination of capacitors, inductors, and in some cases resistors.

29-25. Simple Capacitive Filter

Ripple voltage exists because energy is supplied in pulses to the load by the rectifier. The fluctuations can be reduced considerably if some energy can be stored in a capacitor while the rectifier is delivering its pulse and allowed to discharge from the capacitor between pulses.

Figure 29-30A shows the output of a half-wave rectifier. This pulsating voltage is applied across a filter capacitor (C in Figure 29-30B) to supply the load, R_L. Because the rate of charge of C is limited only by the reactance of the transformer secondary and the forward resistance of the diode in the rectifier, the voltage across the capacitor can rise nearly as fast as the half-sine-wave voltage output from the rectifier. In other words, the RC charge time is relatively short. The capacitor, C, therefore, is charged to the peak voltage of the rectifier within a few cycles. The charge on the capacitor represents a storage of energy. When the rectifier output drops to zero, the

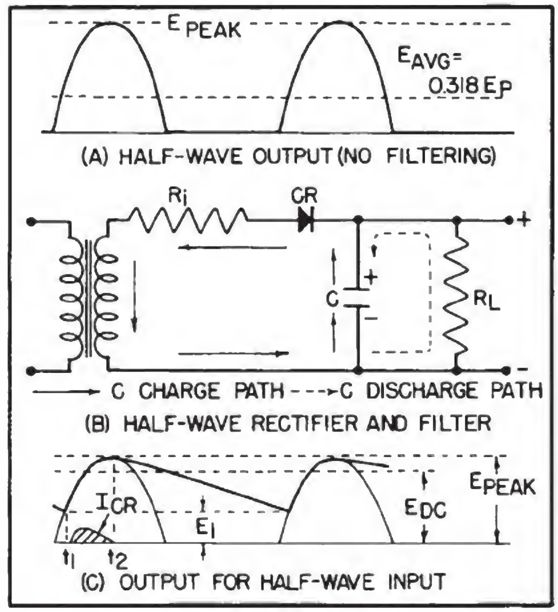

Figure 29-30 - Capacitive filter and waveforms.

voltage across the capacitor does not fall immediately. Instead, the energy stored in the capacitor is discharged through the load during the time that the rectifier is not supplying energy (when the anode is negative). The voltage across the capacitor (and the load) falls off very slowly if it is assumed that a large capacitance and a relatively large value of load resistance are employed. In other words, the RC discharge time is relatively long. The amplitude of the ripple therefore is greatly decreased, as may be seen in Figure 29-30C.

Figure 29-31A shows the input voltage to the filter when a full-wave rectifier is used, and Figure 29-31B shows the resulting output voltage waveform.

After the capacitor has been charged (with either half-wave or full-wave input), the rectifier does not begin to pass current until the output voltage of the rectifier exceeds the voltage across the capacitor. Thus, in Figures 29-30C and 29-31B, current begins to flow in the rectifier when the rectifier output reaches a voltage equal to the capacitor voltage. This occurs at time t_1, when the rectifier output voltage has a magnitude of E_1. Current continues to flow in the rectifier until slightly after the peak of the half-sine wave, at time t_2. At this time the sine-wave voltage is falling faster than the capacitor can discharge. A short pulse of current, beginning at t_1 and ending at t_2, is

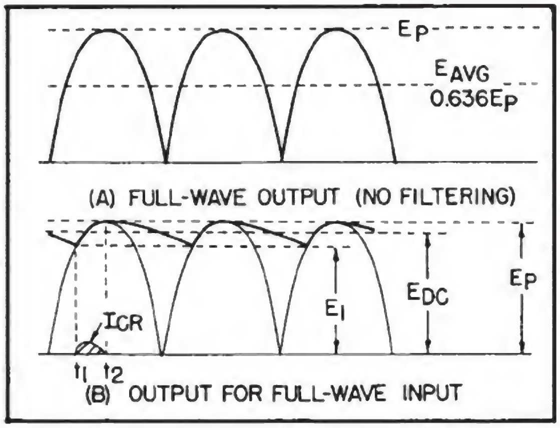

Figure 29-31 - Waveforms for full-wave rectifier capacitive filter.

therefore supplied to the capacitor by the power source.

The average voltage of the rectifier output is shown in Figure 29-30A and 29-31A. Because the capacitor absorbs energy during the pulse and delivers this energy to the load between pulses, the output voltage can never fall to zero. Hence, the average voltage of the filtered output is greater than that of the unfiltered input. However, if the resistance of the load is small, a heavy current is drawn by the load and the average or direct voltage falls. For this reason, the simple capacitor filter is not used with rectifiers that must supply a large load current. Also the input capacitor acts like a short circuit across the rectifier while the capacitor is charging. Because of this high peaked load on the rectifier a current limiting resistor, such as R_1 in Figure 29-30B, is used to protect the diode.

Q24. What would be the effect on circuit operation if the current limiting resistor was installed on the filter side of the diode rather than on the transformer side?

29-26. Pi-Section Filter

The ripple voltage present in a rectifier output cannot be eliminated adequately in many cases by the simple capacitor filter. Filters that are much more effective can be made if both inductors and capacitors are used. The function of the capacitor is to store and release energy, while the inductors simultaneously tend to prevent change in the magnitude of the current. The result of these two actions is to remove the ripple from the rectifier output and to produce a voltage having a nearly constant magnitude.

Figure 29-32 shows a circuit diagram of an inductance-capacitance filter used primarily with receiver power supplies and other low-current power supplies.

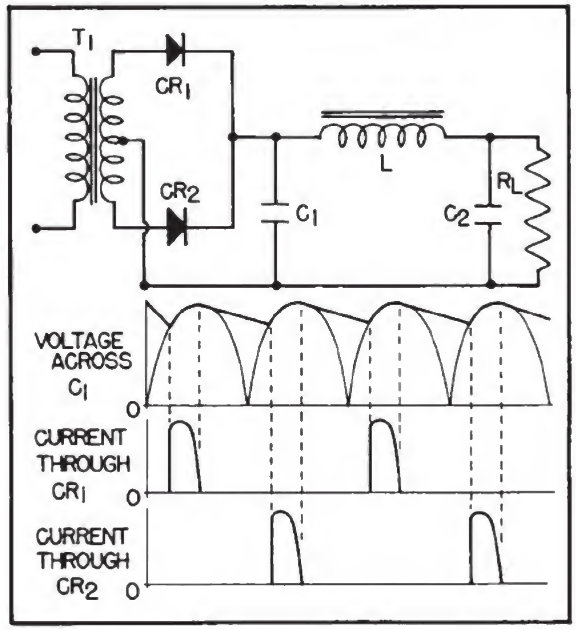

Figure 29-32 - Waveforms of current and voltage in rectifier with pi-section filter.

This type of filter is given the name PI-SECTION because the configuration of the schematic diagram resembles the Greek letter, π . It is also called a CAPACITOR INPUT FILTER. With this type of filter the output waveform closely approximates that of pure direct current. The series choke in the pi-section filter serves to maintain the current at a nearly constant level during the charging and discharging cycle of the input capacitor.

At the bottom of Figure 29-32 are shown the waveforms of current through CR1 and CR2 and the voltage across C1. The current flow through the rectifier diodes is a series of sharp peaked pulses, because the input capacitor acts like a short circuit across the rectifier while the capacitor is charging.

Q25. In a pi-section filter (Figure 29-32) give the discharge path of the input capacitor.

29-27. L-Section Filter

A type of filter used primarily in high-current applications is the L-section filter, so named because of its resemblance to an inverted "L". A schematic diagram of this type of filter is shown in Figure 29-33. The com-

A24. No difference in circuit operation because the resistor is in series with the diode.

A25. Through the load, the series choke and back to the capacitor.

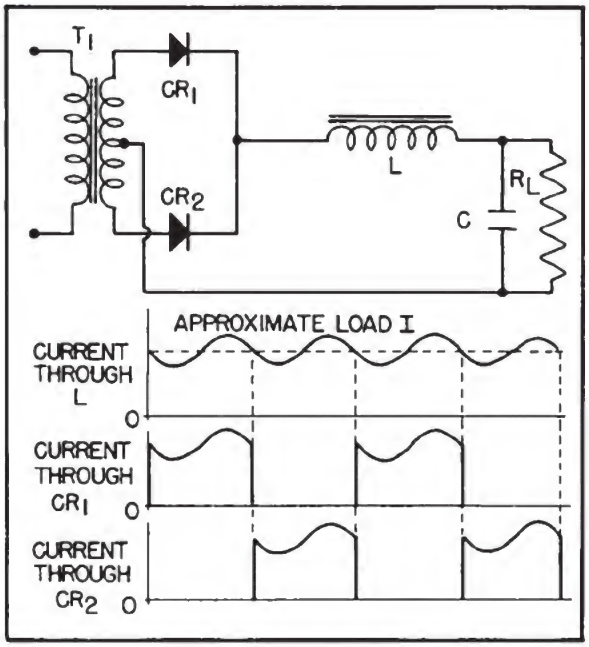

Figure 29-33 - Current and voltage waveforms in full-wave rectifier with L-section filter.

ponents perform the same functions as in the pi-section filter except that the inductor, or choke, input reduces the voltage output of the filter as compared to a capacitive input filter. This filter is also called a CHOKE INPUT FILTER. The input choke allows a continuous flow of current from the rectifier diodes rather than the pulsating current flow demanded by the capacitor input filter. The L-section filter is seldom used with half-wave rectifiers because there is no device to maintain current flow through the load between the half cycles.

Because of the uniform flow of current, the L-section filter has the advantage of better voltage regulation. The inductive reactance of the choke reduces the ripple voltage without reducing the dc output voltage.

Its advantage lies in the fact that it allows each rectifier diode to operate at a relatively constant level of current flow during its half-cycle of operation. This type of operation allows a rectifier to supply the maximum current to the load that it is capable of delivering. A disadvantage of the L-section filter is that instead of delivering a voltage equal to the peak value of the transformer secondary, it supplies a voltage equal to the average of the ac voltage delivered to the rectifier.

Two L-section filters are sometimes used in series to obtain a higher degree of filtering action.

SOLID STATE VOLTAGE REGULATORS

The output voltage developed by any source of power tends to decrease when current is drawn from the source. Most electronic gear used in the Navy can operate satisfactorily with a certain amount of variation in the supply voltage without suffering severe operational deficiency. However, some circuits are very critical and even a slight deviation from the normal supply voltage will cause unsatisfactory operation. These circuits require the use of some type of voltage-regulating device.

29-28. Glow-Tube Voltage Regulator

In a glow-discharge tube, such as the neon glow tube, the voltage across the tube remains constant over a fairly wide range of current through the tube. This property exists because the degree of ionization of the gas in the tube varies with the amount of current that the tube conducts. When a large current is passed, the gas is very highly ionized and the internal impedance of the tube is low. When a small current is passed, the gas is lightly ionized and the internal impedance of the tube is high. Over the operating range of the tube, the product of the current through the tube and the internal impedance of the tube is practically constant.

A simple glow-tube regulator is shown in Figure 29-34. The load current and the current that flows in the neon glow tube both pass through the series resistor, R. If the supply voltage drops, the voltage across the neon tube tends to drop. Therefore, the gas in the neon tube deionizes slightly and less current passes

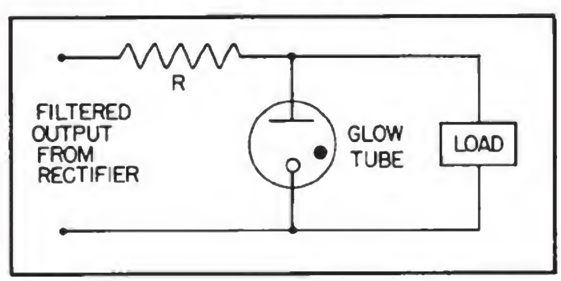

Figure 29-34 - Simple glow-tube voltage regulator.

through the tube. The current through R is decreased by the amount of the current decrease in the tube. Because the current through R is less, the voltage drop across R is less. If the resistor is of the proper value relative to the load and to the glow tube that is used, the voltage across the load is held fairly constant. In any case, the value of R must not be so large that the neon tube fails to ionize.

Glow tubes are designed to operate at various values of voltage ranging from 75 to 150 volts.

29-29. Crystal Diode Voltage Regulator

Equipment using circuits employing solid state devices require power supply voltages and currents much lower than those normally delivered by electron tube power supplies. The value of voltage for a solid state power supply is on the order of 20 volts or less. Even though the voltages are lower there are many instances where voltage regulation is required. The glow-tube regulator is useless in this application because the ionizing potential is far higher than the potentials employed. Therefore, a specially manufactured crystal (semiconductor) diode is required that will operate in the same manner as the glow-tube regulator but at lower potentials. Crystals manufactured for the purpose of voltage regulation are called ZENER DIODES, AVALANCHE DIODES, or BREAKDOWN DIODES. These diodes utilize the breakdown voltage and the avalanche current region of the PN junction, hence their name.

A breakdown, zener, or avalanche diode operated with forward bias will behave in much the same manner as a regular PN junction. In the normal operation of a regular PN junction great care is taken to assure that the breakdown voltage is not exceeded for fear of destroying the crystal. However, the voltage regulator diodes are designed to operate in the avalanche region without damage. The voltage-current characteristics of a typical ZENER Diode is shown in Figure 29-35 for convenience.

The operation of an avalanche diode as a voltage regulator is accomplished in the following manner. With small reverse bias across the PN junction the barrier potential is increased. Only a small leakage current will flow because the majority current carriers in both P and N sides of the junction are attracted away from the junction. This action leaves a space charge depletion layer adjacent to the junction.

Increasing the reverse voltage across the junction increases the velocity of the minority carriers in this region. (The intensity of the junction field is increased and hence increased acceleration of the minority carriers). Some

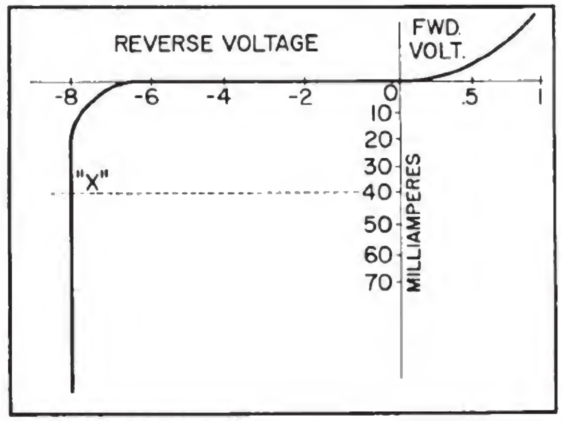

Figure 29-35 - Volt-current characteristics of a typical zener diode IN63(25°C).

of these carriers collide with covalent bond electrons releasing them as carriers. This action has a cumulative effect called avalanche ionization. It comprises a rapidly rising reverse current that, unless checked by a series limiting resistor, may destroy the semiconductor. The reverse voltage at which avalanche effect occurs is called the REVERSE BREAKDOWN VOLTAGE and is abbreviated BV_R.

The symbol B is used in the crystal diode (CR, Figure 29-36) to indicate operation in the reverse breakdown voltage region. When operated in this way the voltage across the load is held constant because the load is in parallel with the crystal diode, and the reverse breakdown voltage of the diode is approximately constant for wide changes in diode current.

If the supply voltage decreases (Figure 29-36) the reverse voltage across the semiconductor diode CR will tend to decrease. Thus the speed of the current carriers in the crystal will decrease and the number of collisions with valence electrons will decrease with the result that the reverse current through the crystal will decrease. The current through R is decreased by the amount of the current decrease through crystal CR. Because the current through R is

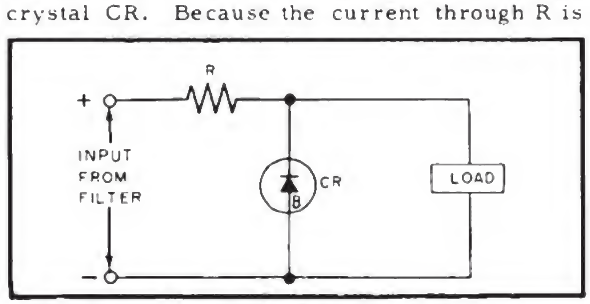

Figure 29-36 - Breakdown diode voltage regulator.

decreased, the voltage drop across R is proportionately less. If R has the right value relatively to the load and to CR, the voltage across the load will remain fairly constant. However, R must not be so large that the avalanche effect will not take place in the crystal.

If a source voltage of 12 volts is applied to the input of Figure 29-36, and the desired load voltage is 8 volts (as indicated by point "x" in Figure 29-35) the voltage drop across R must be equal to 4 volts. e.g., 12V - 4V = 8V. Assuming a load current of 40ma and a Zener reverse current of 40 ma, the current flowing through R would be equal to 80ma. The value of R can be determined by use of Ohm's law as follows:

$$R = \frac{E}{I}$$

$$R = \frac{4V}{80ma}$$

$$R = 50 \text{ ohms}$$

Zener diodes are designated to operate at various voltages. When a regulated voltage in excess of the rating of one Zener diode is required, two or more diodes may be connected in series. This arrangement permits several regulated voltages with small current drain to be obtained from a single rectifier power supply. The breakdown diode, when operated in the forward biased direction, has a negative temperature coefficient. That is, an increase in ambient temperature will cause a decrease in junction resistance. When operated in the avalanche region the breakdown diode will exhibit a positive temperature coefficient (increased ambient temperature will result in increased junction resistance.

EXERCISE 29

1. What is the name of the general class of materials used in the manufacture of PN junctions?
2. Name two materials from which PN junctions are made.
3. What is the function of the electrons in the outermost orbits of an atom?
4. How many orbital electrons does the element gallium have if its atomic number is 31?
5. An element has an atomic number of 34. What is its outer most subshell and how many electrons does this subshell possess?
6. Define the terms: valence shell, valence electron, and valence number. Give the valence number of silicon.
7. Describe an ionic bond.
8. Describe a covalent bond.
9. What is intrinsic material?
10. What is the difference between a valence band electron and a conduction band electron?
11. How may valence bonds be broken in semiconductor crystals?
12. What phenomenon occurs when heat, at room temperature, is applied to an intrinsic semiconductor material?
13. What is a donor atom?
14. What is a donor ion?
15. What is an acceptor atom?
16. What is an acceptor ion?
17. What is meant when a crystal is said to be "doped"?
18. Describe P-type material.
19. Describe N-type material.
20. Define the term "hole" in your own words.
21. To what degree is a crystal doped when a PN junction is formed?
22. Explain how it is possible for an electron from a donor atom to exist in the forbidden region of a silicon crystal.
23. Describe, in your own words, how a hole is created.
24. Since an N-type material contains excess electrons, what is the charge on an isolated piece of N material? Explain your answer.
25. Explain why a PN junction possesses rectifying properties.
26. By what process is the junction field formed?
27. When is equilibrium established in a PN junction.
28. What is the net charge on a PN junction? Explain your answer.
29. In order to obtain a positive output from a half-wave rectifier would the anode or the cathode of the PN junction be connected to the transformer winding? Always?
30. What type of circuit is shown in Figure 29-37. Will it operate? If so, explain its operation giving current paths and waveforms, etc.

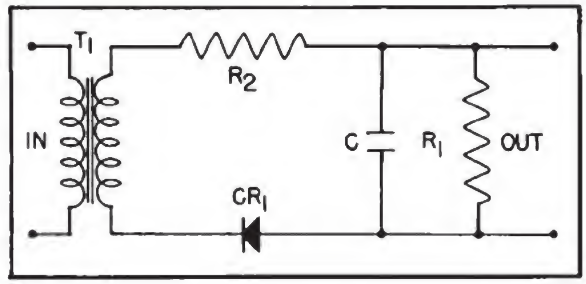

Figure 29-37

31. Sketch a bridge rectifier circuit with a Pi-section filter that will provide a negative output voltage. Show current paths, etc.
32. Why would a capacitive input filter and a half-wave rectifier be unsuitable for use as a power supply when the load resistance is very small?
33. If current were the prime consideration of a power supply, rather than voltage, what type of filter would most likely be used?
34. Describe the function of the series resistor in a simple voltage regulator.
35. What happens to the majority current in a breakdown diode when the reverse bias is increased?
36. It is desired to regulate the output voltage of the power supply in Figure 29-38. Sketch your idea of how a breakdown diode voltage regulator and load might be connected and explain how your regulator would work (include polarities, current paths, values, etc.).

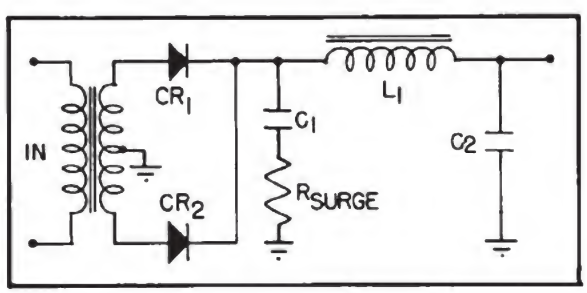

Figure 29-38

CHAPTER 30

TWO JUNCTION TRANSISTORS

The transistor is a relatively new form of electronic device. It can perform many of the functions of an electron tube, and in addition, can do some things better and more efficiently. Electron tubes depend on the flow of electrons through a vacuum, a gas, or a vapor; whereas, transistors make use of the flow of current carriers through a special type of solid known as a SEMICONDUCTOR.

Although semiconductor diodes can permit more current to flow in one direction than the other (ability to rectify), they cannot amplify a signal. Three element semiconductors (like the three element electron tube) are needed in order to amplify a signal.

Semiconductors that can amplify a signal are called TRANSISTORS. There are many different types of transistors with individual characteristics, but the theory of operation is basic to all of them.

The advent of the transistor has opened a completely new field for the development of portable equipment. The compactness and ruggedness of transistorized equipment has allowed the manufacture of portable equipment that was previously impractical, due to weight and expense, etc.

Transistors are now being used in mobile equipment, test equipment, tape recorders, photographic equipment, hearing aids, radios; and the list goes on without end. In other words, transistors may be used in almost any application where low and medium power electron tubes are used.

30-1. Comparison of Electron Triode and Two Junction Transistors

In some ways the triode electron tube and the junction transistor are similar. Both have three elements, and these three elements may be compared in the way they work to produce amplification.

In the electron tube triode the three elements are:

1. The CATHODE which gives off or emits electrons.
2. The GRID which controls the flow of electrons.
3. The PLATE which attracts or COLLECTS the electrons.

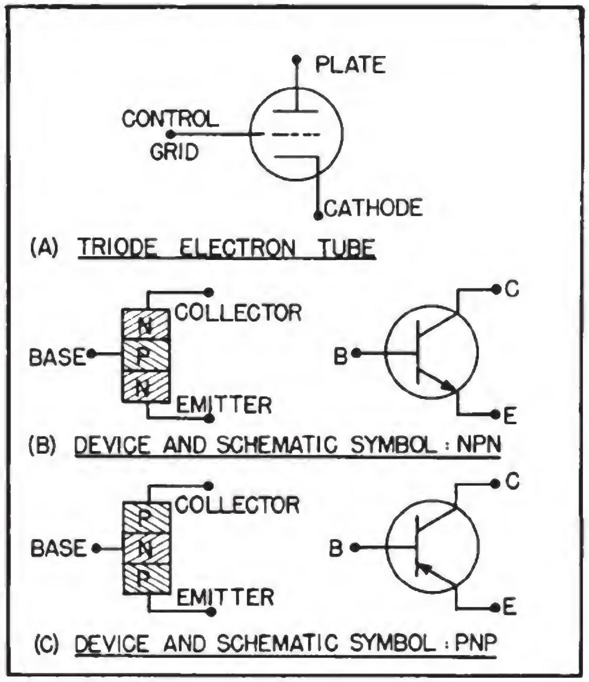

Figure 30-1 - Corresponding elements of triode electron tube and junction transistor.

The three elements in the two junction transistor are:

1. The EMITTER which gives off, or EMITS current carriers (electrons or holes).
2. The BASE which can control the flow of the current carriers.
3. The COLLECTOR which collects the current carriers emitted by the emitter.

Transistors are classed as PNP or NPN according to the arrangement of the impurities in the crystal. Transistors contain two of the basic PN junctions previously described. One PN junction is between the emitter and the base. The other PN junction is between the collector and the base.

The three elements in a triode electron tube

and the corresponding elements in the PNP and NPN junction transistors are illustrated in Figure 30-1. The corresponding schematic symbols for the PNP and NPN junction transistors are shown at the right of the device symbols in Figure 30-1(B)(C).

In the PNP transistor, the collector collects holes. In the NPN transistor the collector collects electrons. The schematic symbol for both types is similar but not identical. A heavy straight line represents the base and two slanting lines to the base represent the emitter and collector terminals. An arrowhead is placed on the emitter line and points toward the base for PNP transistors, and away from the base for NPN transistors. The ARROWHEAD ALWAYS POINTS TOWARD THE N-TYPE MATERIAL. Thus, if the arrowhead points toward the base line the base is N type material and the transistor is a PNP transistor. If the arrowhead points away from the base line (toward the emitter) the emitter material must be N type. Since the base is always of the opposite type material, the base must be of a P type material and the transistor is an NPN transistor.

30-2. Biasing and Basic Current Paths

It has been mentioned that the two junction transistor contains two PN junctions. One of these junctions exists between the emitter material and the base material, and is called EMITTER-BASE JUNCTION. The other junction exists between the collector material and the base material and is called the COLLECTOR-BASE JUNCTION.

As mentioned in Chapter 29, each PN junction contains a depletion region. This depletion region is formed initially by the diffusion and recombination of majority carriers in the vicinity of the junction. During the diffusion process, positive ions in the N region and negative ions in the P region are left uncompensated (the mobile carriers which balanced them are lost in recombination). These charged atoms, or ions as they are called, are fixed in place in the crystal lattice structure, and cannot move. Thus, they make up a layer of fixed charges on the two sides of the junction. On the N side of the junction there is a layer of positively charged ions; on the P side of the junction there is a layer of negatively charged ions. Therefore, an electric field is established as shown in Figure 30-2. In the unbiased junction, the magnitude of this field is such that it exactly neutralizes the tendency for diffusion, and the major portion of the recombination action ceases.

The depletion region is an area where few free carriers exist. The reason for this is the junction field. A free carrier coming under the influence of the electrostatic field will be accelerated out of the field (electrons are accelerated against the arrow and holes are accelerated with the arrow). After the junction field is established, the free electrons in the N type material must overcome the energy of the field (traveling with the arrows) to reach the P type material.

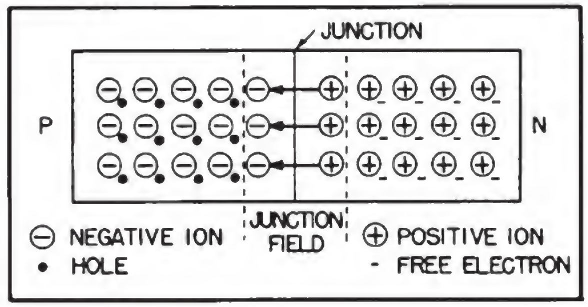

Figure 30-2 - Electrostatic field of PN junction.

Application of bias to a junction causes the junction field to increase or decrease. Application of forward bias, as in Figure 30-3, causes the junction field to be reduced; thereby making it easier for the free electrons to cross the junction. Increasing the forward bias further decreases the opposition of the junction field, thereby, increasing current flow through the junction.

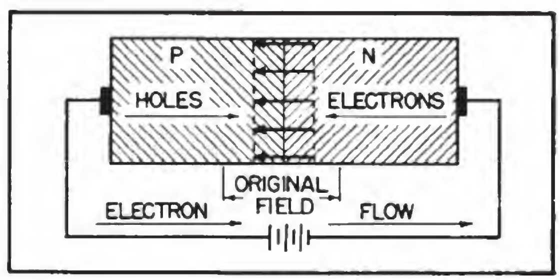

Figure 30-3 - Forward biased PN junction.

Application of reverse bias, as in Figure 30-4 causes the junction field to be increased; thereby; making it more difficult for the majority carriers to cross the junction. Increasing the reverse bias further increases the magnitude of the junction field, and thus, increases the opposition to majority current flow.

Although semiconductor diodes can permit more current to flow in one direction than the other (ability to rectify), they cannot amplify a signal. Three element semiconductors are needed in order to amplify.

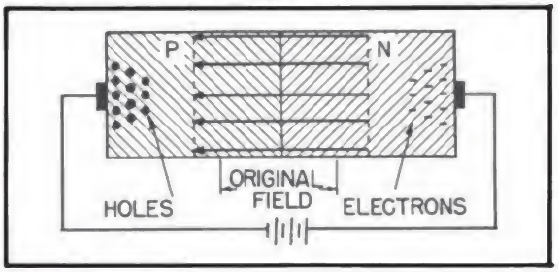

Figure 30-4 - Reversed biased PN junction.

One type of junction transistor is formed by introducing a thin region of P type material between two regions of N type material in a single crystal of silicon or germanium. The transistor so formed is called an NPN transistor. By introducing a thin region of N type material between two regions of P type material a PNP transistor is formed.

The electrostatic junction fields are established in the same manner as in the basic PN junction. Figure 30-5 shows a pictorial view of an unbiased NPN transistor. The ions and mobile carriers are included along with the junction fields in order to indicate the charge distribution in the transistor.

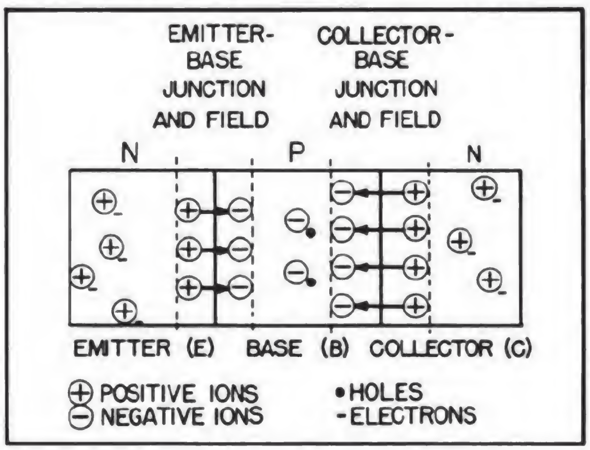

Figure 30-5 - Unbiased two junction transistor.

In this unbiased transistor, diffusion and recombination has already occurred and the junction fields have been established. The base in Figure 30-5 is formed of P type material and contains negative ions (due to acceptor atoms) and holes (due to ionized germanium or silicon atoms). The emitter and collector are both N type material and contain positive ions (due to donor atoms) and free electrons.

The collector-base junction field is larger and stronger than the emitter-base field. This is accomplished by controlling the amount of impurities (doping) added to each element. Because the doping of each element is highly variable, depending on the job the transistor is designed to do, no attempt will be made to explain the doping procedure.

In normal operation the emitter-base junction is forward biased and the collector-base junction is reverse biased. Forward biasing of the emitter base junction reduces the size and intensity of the emitter-base field. Therefore, the emitter-base junction is usually referred to as the low resistance junction. Reverse biasing the collector-base junction increases the size of the collector-base field. Reverse biasing opposes conduction by majority carriers, thus, the collector-base junction is referred to as the high resistance junction. Figure 30-6 is a pictorial diagram of the two junction transistor showing the effect on the junction fields when the two biasing batteries are connected. The bias batteries are labeled V_{CC} for the collector voltage supply and V_{EE} for the emitter voltage supply. In transistors the letter V is used in place of the letter E to indicate voltage due to the fact that the letter E is used to indicate the emitter electrode. A list of common symbols used in connection with semiconductor devices is included at the end of this chapter.

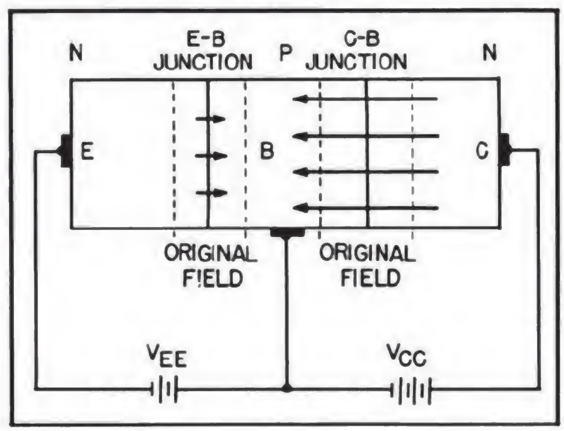

Figure 30-6 - Biased two junction transistor.

If the transistor is of the PNP type the bias batteries and the direction of the junction fields will be reversed.

The basic current paths through an NPN transistor are illustrated in Figure 30-7A. The solid arrows indicate current flow due to free electrons. The dotted arrows indicate current flow due to hole movement (valence band electron movement). Current flow in the external circuit is always due to the movement of free electrons. Emitter current (I_E) is shown leaving the emitter supply battery and flowing to the N type emitter. Since electrons are the majority carrier in N material the electrons will move through the emitter to the emitter-base field. Having gained energy from V_{EE} the

electrons will overcome the small opposition of the forward biased junction and be injected into the base region. However, the base region is P type material and the electrons are now MINORITY carriers. As they move into the base region, some of them will recombine with available holes. The electrons that recombine will move out through the base lead as base current (I_B), and return to the emitter supply battery V_{EE}. Most of the electrons that move into the base region will eventually come under the influence of the collector-base junction field. The direction of the reverse biased collector-base field is such that minority carriers will be accelerated. Since the electrons are minority carriers at this time, they will be accelerated through the field and injected into the collector region. The collector is N-type material and the electrons are again majority current carriers. The electrons will move through the collector material to the collector terminal. They will then move out of the collector and return to the positive terminal of V_{CC} as collector current (I_C).

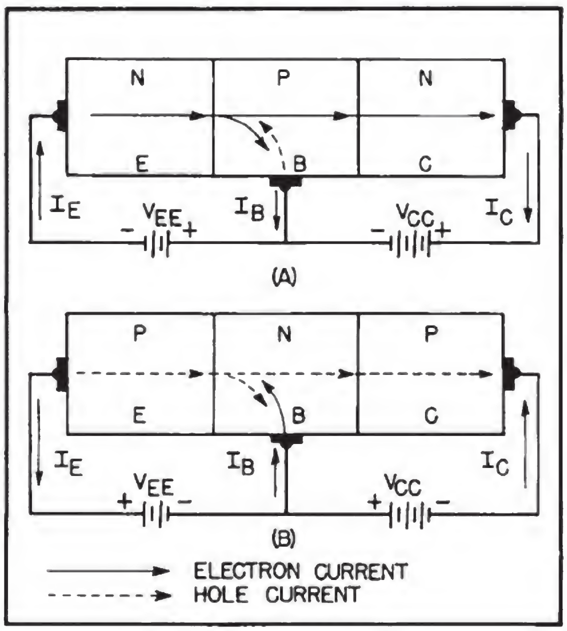

Figure 30-7 - Basic current paths for NPN and PNP transistors.

The basic current paths in a PNP transistor are shown in Figure 30-7B. The majority current carrier in the PNP transistor is the hole. Briefly, the method of transporting a current through a PNP transistor is as follows: The energy supplied by the positive terminal of the emitter supply (V_{EE}) causes an electron hole pair to be generated at the emitter terminal. The electron travels through the wire to the positive terminal of V_{EE} as emitter current, I_E. The hole travels through the transistor from emitter to collector, by the method previously described in Chapter 29, as hole movement. Upon reaching the collector terminal the hole will recombine with an electron from the negative terminal of the collector supply (V_{CC}).

While the current flow in the external circuit of the PNP transistor is opposite to that of the NPN, it will be noticed that, regardless of the type of transistor, the MAJORITY CARRIER always moves through the transistor FROM EMITTER TO COLLECTOR.

Some holes, which are injected into the base by the emitter, will not complete the journey to the collector, but instead will recombine with electrons in the base. This recombination causes electrons to be drawn from the negative terminal of V_{EE}, through the base lead, and into the N type base material. The result is a small base lead current.

Although there are other factors (discussed in the next section) involved in the transportation of a current through a transistor, the basic action is as just described.

Q1. Why are an electron tube cathode and a transistor emitter similar?

Q2. In biasing a PNP transistor would the negative or positive terminal of V_{EE} be connected to the emitter?

Q3. If the base of a PNP transistor were more positive than the collector what type of bias would the collector-base junction have?

30-3. Drift, Diffusion and Base Recombination

DRIFT, DIFFUSION and RECOMBINATION in the base region are factors directly affecting the movement of current through a junction transistor. Therefore, before going further, a brief discussion of these factors will be of value.

Whenever a current carrier (such as an electron) comes under the influence of an electric field, it will drift in a predetermined direction. For instance, if a potential difference is applied to a piece of copper wire, current will flow through the wire from the negative potential toward the positive potential. AT ANY GIVEN INSTANT many of the electrons in the wire will be moving in directions tangent or even opposite to the direction of the main current flow. In other words, the electrons will not move in a direct line from the negative to the positive potential but will take a wandering path. If it were possible to look at a few electrons moving

through the wire, their paths might appear something like that in Figure 30-8. The important point to remember is that when an electric field is established in a material which has an abundance of current carriers, more of the carriers will move in a direction determined by the field than will move in other directions. By way of definition: DRIFT CURRENT is the flow of charge carriers under the influence of an electric field.

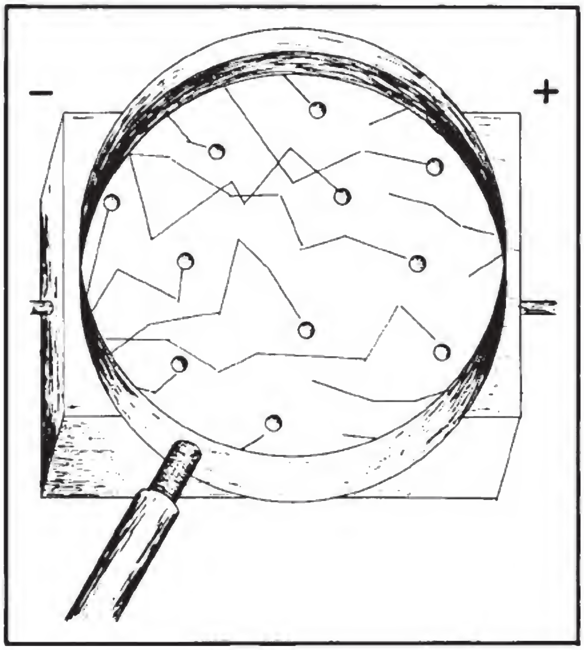

Figure 30-8 - Drift superimposed on random path of single carrier.

A charge flow will also occur if there is a difference in DENSITY of carriers between one area and another of a given volume of material. In other words, if more carriers exist at one side of a piece of semiconductor material than exist at the other, there will be a charge flow from the side with the greater density to the side with the lesser density. When no external influences are applied, the carriers (electrons or holes) will distribute themselves evenly throughout the crystal. This is analogous to the action which occurs when smoke is blown into a sealed glass through a small hole. Given a small amount of time the smoke will distribute itself evenly throughout the volume of the glass. Charge flow that takes place under these conditions is called DIFFUSION CURRENT. As before, each individual carrier has an unpredictable random motion due to thermal energy, but when there is a difference in carrier concentration (density gradient), MORE CARRIERS WILL MOVE TOWARD THE AREA OF LESSER DENSITY THAN WILL MOVE IN ANY OTHER DIRECTION. By way of definition: DIFFUSION is a charge flow due to a density gradient and not an electric field.

Carriers that enter the base region and recombine are lost as far as collector current is concerned. Thus, for the transistor to be efficient, as many of the emitter carriers as possible must reach the collector region. If the base region is too wide, all the carriers injected into the base by the emitter will recombine and leave the crystal as base current, (I_B). The solution for this problem is to make the base region very thin, thereby reducing the opportunity for a carrier to recombine and be lost. Figure 30-9 shows a realistic pictorial diagram of a transistor which illustrates the thin base region.

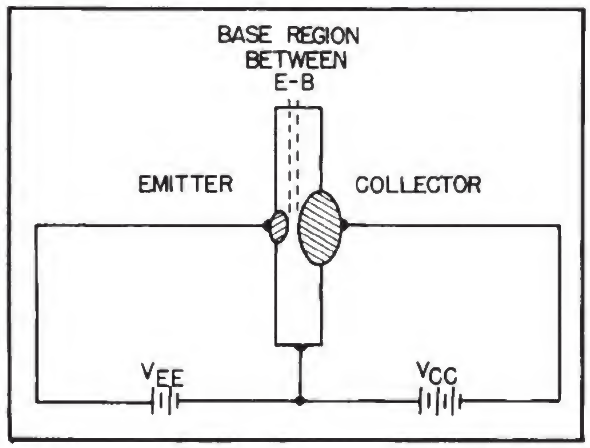

Figure 30-9 - Pictorial diagram of a PNP transistor.

The collector is made physically larger, first to increase the chance of collecting carriers that diffuse to the side as well as directly across the base region; and second to give it a larger heat dissipation factor (explained later). In practical transistors 92% to 99% (and in many cases even more) of the carriers emitted by the emitter reach the collector.

Q4. If a semiconductor material had a density gradient (concentration of carriers in one area) plus an electric field, would it be possible to have diffusion current and drift current occur at the same time?

Q5. What happens to the 1% to 8% of emitter carriers that do not reach the collector?

A1. Both elements emit the major current carrier for their respective devices.

A2. Positive.

A3. Reverse bias.

A4. Yes. The result would be drift current superimposed on diffusion current.

A5. They leave the crystal as base current (I_B).

30-4. Base Lead Current

In Figure 30-7A the electron current (I_B) in the base lead for an NPN transistor is shown to flow away from the base region. This may, or may not, be the case for an NPN transistor used in a specific application. Figure 30-7A neglects the REVERSE CURRENT flow between the collector and base. In other words, if the emitter lead were opened, there would be a very small reverse current flow between the collector and base. This current flow is called I_{CBO}. The letter "I" indicates current, the subscripts "C" and "B" indicate the collector and base electrodes; and the subscript "O" means emitter open. Thus, I_{CBO} means collector to base current measured with the emitter open. Figure 30-10 shows an NPN transistor with the emitter circuit open and only I_{CBO} flowing.

NOTE: The block pictorial diagram is used for ease of demonstration.

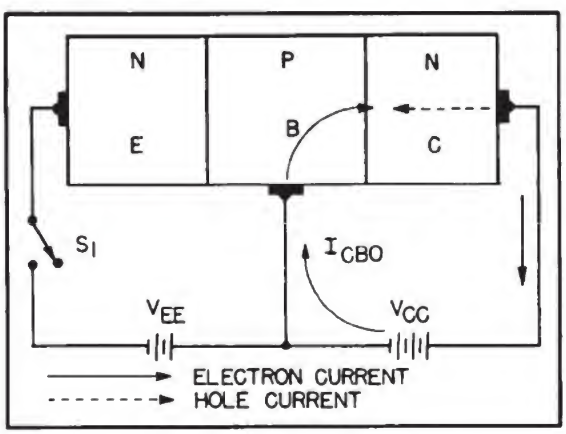

Figure 30-10 - I_{CBO} in an NPN transistor.

The flow of I_{CBO} is opposite to the normal base lead current (I_B). However, under normal operating conditions the magnitude of the reverse current is too small (in the microampere range) to subtract a significant amount from the base current. Under normal conditions the base lead current (I_B) for an NPN transistor will flow AWAY from the base and for a PNP, base current (I_B) will flow TOWARD the base, although it should be remembered that the base lead current for either type of transistor depends on the percentage of emitter current that recombines in the base region; and the magnitude of the reverse current (I_{CBO}). I_{CBO} increases rapidly as temperature increases (exponential increase). I_{CBO} doubles in value for every 8 to 10 degrees centigrade rise in temperature.

Q6. Why does I_{CBO} in an NPN transistor consist of electron flow in the P type material and hole current in the N type material?

30-5. Current Gain in the Basic Transistor

Current in the emitter circuit can be varied by changing the emitter-base voltage. The change in the emitter current will cause an almost equal change in the collector current. The change in the collector current is always slightly less than the change in the emitter current because some of the emitter current will leave the transistor as base current. The collector current is the difference between the emitter current and the base current, or:

$$I_C = I_E - I_B \qquad (30\text{-}1)$$

The current gain of a transistor is analogous to the amplification factor of a triode electron tube. The transistor, however, has either of two current gains depending upon the circuit connection.

Transistor current gain in one type of circuit connection is the ratio of the change in collector current to the corresponding change in emitter current for a constant collector to base voltage. In this circuit, current gain is designated as alpha (α). The equation is:

$$\alpha = \left.\frac{\Delta I_C}{\Delta I_E}\right|_{V_{CB}\ \text{constant}} \qquad (30\text{-}2)$$

Transistor current gain for another type of connection is the ratio of the change in collector current to the corresponding change in base current for a constant collector to emitter voltage. In this circuit, current gain is designated as B (pronounced beta). The equation is:

$$B = \left.\frac{\Delta I_C}{\Delta I_B}\right|_{V_{CE}\ \text{constant}} \qquad (30\text{-}3)$$

The method for solving for current gain can best be expressed by use of an example.

With collector voltage adjusted to -10 volts and base current equal to 25 microamperes, collector current I_C=1 ma or 1000 ua. Increasing I_B to 125 ua (with constant V_c) increases

the collector current to approximately 5 ma or 5000 ua. Thus, an increase in input base current of 125-25 = 100 ua causes an increase in collector current of 5000-1000 = 4000 ua. The current gain is:

$$B = \text{beta} = \frac{\Delta I_C}{\Delta I_B} = \frac{4000}{100} = 40$$

Most transistor manuals give either the alpha or beta figures. If it is desired to determine either one of these quantities when the other is known, the following equations may be used:

$$B = \frac{\alpha}{1 - \alpha} \qquad (30\text{-}4)$$

$$\alpha = \frac{B}{1 + B} \qquad (30\text{-}5)$$

where: B = beta current gain (always greater than one)

α = alpha current gain (always less than one)

For instance, a transistor has an alpha of 0. 97, determine the beta by use of equation (30-4).

$$B = \frac{\alpha}{1 - \alpha} \qquad (30\text{-}4)$$

$$B = \frac{0.97}{1 - 0.97}$$

$$B = \frac{0.97}{0.03}$$

$$B = 32.33$$

Q7. A transistor is operating under the following conditions: V_C = 12 V, I_C = 1. 5 ma, I_E = 1. 53 ma. Increasing I_E to 4. 59 ma causes the I_C to increase to 4. 5 ma. (V_{CB} held constant at 12 V). Determine the alpha of the transistor.

30-6. Basic Amplifier

Transistor triodes, like electron tube triodes, are amplifiers. The current, voltage, and power gain depend on the alpha or beta values and the ratio of the output to input resistance. Although electron tube circuits are similar, they have widely differing characteristics. It is not possible to substitute a transistor for an electron tube without changing bias and plate supply voltages. For example, the transistor in some circuits is a current amplifier, and has a low input resistance; the corresponding electron tube amplifier has a high input impedance. The elements of a basic triode electron tube amplifier and the corresponding elements of a basic PNP and NPN junction transistor amplifier are illustrated in Figure 30-11.

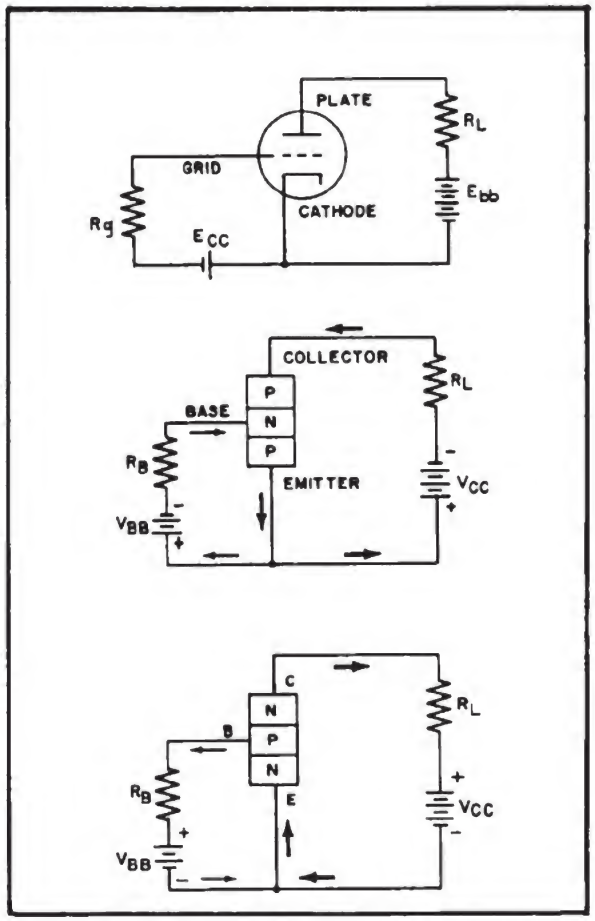

Figure 30-11 - Corresponding elements in triode and transistor amplifiers.

Note that the direction of current flow and the polarities of the biasing batteries are opposite in the two transistor amplifiers. The base supply is designated by the symbol V_{BB}. The collector supply is designated by the symbol V_{CC}.

BASIC AMPLIFIER CONFIGURATIONS

The transistor amplifier may be connected in any one of three basic configurations. These circuits are the (1) COMMON-EMITTER, (2) COMMON-BASE, and (3) COMMON-COLLECTOR. These three arrangements correspond respectively to three electron tube basic

A6. Because I_{CBO} is reverse current, it flows by MINORITY carriers.

A7. Alpha equals 0.98.

$$= \frac{\Delta I_C}{\Delta I_E} \; V_C \text{ held constant}$$

$$= \frac{3}{3.06} = 0.98$$

---

circuits: (1) grounded-cathode, (2) grounded-grid, and (3) grounded-plate amplifier. The significance of the expression "grounded" is that the element which is said to be grounded is, in reality, common to the input and output circuits; and does not necessarily have to be grounded to provide satisfactory operation.

30-7. Common-Emitter Configuration

The widely used grounded-cathode electron tube amplifier and the corresponding common-emitter (sometimes called grounded-emitter) transistor amplifier are illustrated in Figure 30-12. The transistor bias polarities are established for a PNP junction transistor in Figure 30-12B, and for an NPN junction transistor in Figure 30-12C.

The input signal to the triode (Figure 30-12A) is developed across the grid resistor in series with the bias voltage between the grid and cathode. The output signal of the triode is developed between the plate and ground. The average grid voltage depends upon the magnitude of the grid bias supply voltage, and the average plate voltage depends upon the magnitude of the plate supply voltage.

The input signal to the PNP transistor (Figure 30-12B) is developed between the base and emitter in series with the bias voltage in this circuit. The bias polarity is in the forward, or low resistance, direction of the base-emitter junction. The output signal of the transistor is developed between the collector and emitter terminals. The average base voltage, as measured with respect to the emitter, depends upon the magnitude of the bias voltage in the base-emitter circuit. The magnitude of the bias current determines the class of operation of the transistor. The average collector voltage depends upon the magnitude of the collector voltage supply. In many instances, the magnitude of the collector voltage has only a small effect in determining the collector current. THE COLLECTOR CURRENT IS PRIMARILY A FUNCTION OF THE BIAS CURRENT IN THE INPUT CIRCUIT.

When a signal is applied to the input circuit

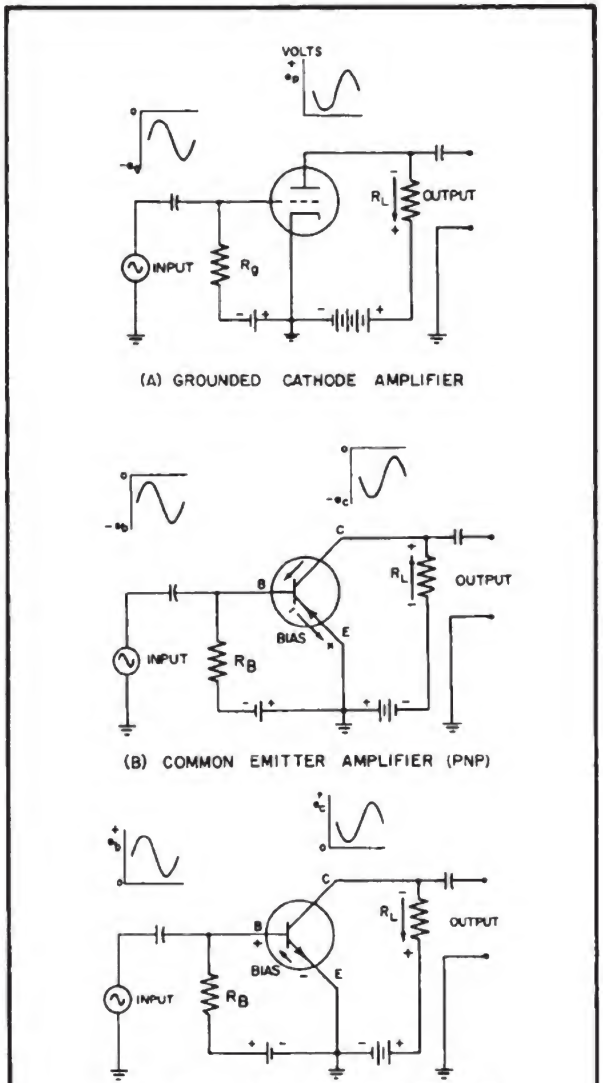

Figure 30-12 - Corresponding electron tube and transistor amplifiers.

of the transistor, the bias current varies about an average, or no-signal, value. This action causes the collector circuit current to vary by a much greater amplitude through the load impedance, which is connected in series with the collector voltage supply and the collector terminal of the transistor.

A phase shift of 180 degrees occurs between the input and output signals. Thus, a positive-going input signal (Figure 30-12B) opposes the base-emitter bias and decreases the base-emitter current. This action decreases the collector current and voltage drop across the

load impedance, resulting in an increase in collector-to-ground output voltage. Because the collector is negative with respect to ground, a positive-going input signal will result in a negative-going output signal.

A positive-going input signal (Figure 30-12C) will cause an increase in base-emitter current with a corresponding increase in collector current and voltage drop across the load impedance. Because this drop subtracts from the collector voltage, the output signal is less positive and is therefore equivalent to a negative-going output signal. The latter action is similar to that occurring in the plate circuit of the triode (Figure 30-12A) with a positive-going signal on the grid. The increase in plate current causes an increased voltage drop in the plate load impedance with a resulting decrease in the plate-to-ground voltage. A decrease in the positive plate voltage is equivalent to a negative-going output signal.

The common-emitter configuration has a medium range of input impedance. For junction transistors, this value may be on the order of 1 k ohm. The output impedance of the grounded-emitter configuration may be on the order of 50 k ohms.

Q8. Would doubling the collector voltage normally be expected to double the collector current?

30-8. Common-Base Configuration

The common-base (or grounded-base) transistor amplifier is analogous to the grounded-grid electron tube amplifier. These circuits are illustrated in simplified form in Figure 30-13. The grounded-grid in the triode (Figure 30-13A) is common to both the input and the output circuits. The common-base in the PNP junction transistor (Figure 30-13B) is common to both input and output circuits.

The input signal to the triode is applied between the cathode and ground and varies the voltage across the cathode resistor in series with the bias voltage source in the grid cathode circuit. The output signal is developed between plate and ground as a result of the variations in plate current through the plate load impedance. Effectively, the input signal is developed between the cathode and the grid, and the output signal is developed between the plate and the grid.. Hence, the grounded-grid forms the common element between the input and output circuits. Plate current flows through the cathode resistor across which the input signal is developed. The action is degenerative and lowers the gain of the amplifier compared to that of a grounded-cathode amplifier.

The input signal to the transistor amplifier is applied between the emitter contact, E, and ground and appears across the resistor in series with the emitter-base bias voltage. The average value of the emitter-to-base voltage depends upon the magnitude of the bias voltage in the input circuit. The output signal voltage is developed between the collector and grounded base. The average value of the collector-to-ground voltage depends upon the magnitude of the collector supply voltage. The base of the transistor is common to both input and output circuits and SHOULD NOT NECESSARILY BE GROUNDED, PROVIDED THE CONTINUITY OF THE COMMON CONNECTION IS MAINTAINED.

The input circuit bias polarity is in the forward, or low resistance, direction of the

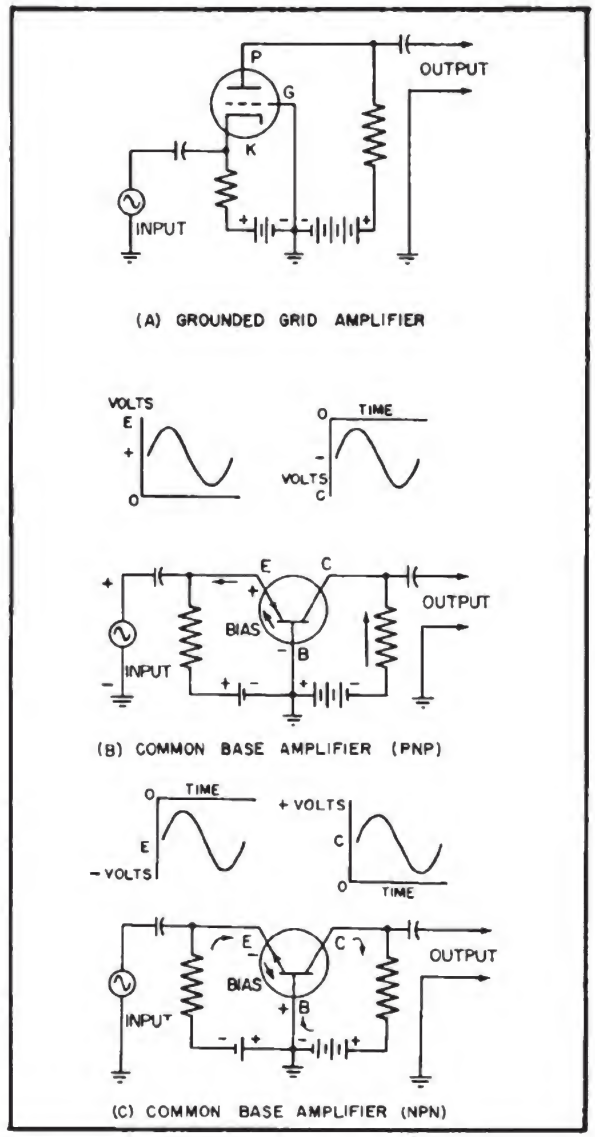

Figure 30-13 - Grounded-grid triode and corresponding common-base transistor amplifiers.

A8. No. V_C has very little effect on I_C.

---

emitter-base junction. As in the case of the common-emitter configuration, the magnitude of the bias current determines the mode of operation of the transistor amplifier, and the collector current is primarily a function of the bias current in the input circuit.

In contrast with the grounded-emitter amplifier, NO PHASE SHIFT OCCURS in the grounded-base amplifier between the input and the output signals. For example, a positive-going input signal (Figure 30-13B) will aid the emitter-base bias and increase the magnitude of the emitter current accordingly. This action will increase both the collector current and the voltage drop across the collector load impedance with a consequent decrease in collector-to-ground voltage. The collector is negative with respect to ground; hence, the decrease in negative voltage to ground is equivalent to a positive-going output signal.

In the NPN transistor amplifier of Figure 30-13C, a positive-going input signal will oppose the input bias and reduce the emitter current accordingly. This action will reduce both the collector current and the voltage drop across the collector load impedance with a resulting increase in collector voltage. Because the collector is positive with respect to ground, the increase in positive voltage is equivalent to a positive-going signal. Thus a positive-going input signal will cause a positive-going output signal in both PNP and NPN grounded-base amplifiers.

The input impedance of a grounded-base transistor amplifier is on the order of 100 ohms or less. The output impedance is relatively high—that is, approximately 500 k ohms for the junction transistor.

Q9. The alpha and beta gains are the ratio of a change in output current to a change in input current. Would alpha or beta be used to describe the gain of a transistor used in the common-base configuration?

30-9. Common-Collector Configuration

The common-collector (sometimes called the grounded-collector or emitter follower) transistor amplifier corresponds to the grounded-plate electron tube amplifier, or cathode follower. These circuits are illustrated in simplified form in Figure 30-14. In the cathode follower the plate is at ground potential with respect to the signal component, and in the corresponding transistor circuits the collector is at ground potential with respect to the associated signal component. In the cathode follower the input signal is applied between the grid and the grounded side of the plate circuit, and the output signal is developed between the cathode and grounded side of the plate circuit. Thus, the plate is common to the input and the output circuits.

In the common-collector transistor amplifier, the input signal is applied between the base and the grounded side of the collector circuit. The output signal is developed between the emitter and the grounded side of the collector circuit. Thus, the collector is common to the input and the output circuits.

Bias current is supplied by the single-cell source in series with the base-emitter circuit, and collector voltage is obtained from a battery consisting of several cells in series between the collector and ground. The base-emitter circuits are biased in the forward, or low resistance, direction, and the collector-base circuits are biased in the reverse, or high resistance, direction. The input impedance of the common-collector transistor is relatively high, because the input circuit includes the degenerative action caused by the emitter load resistor. In effect, the output signal is arranged series opposing to the signal source, thereby reducing the signal current it must supply. The output impedance is relatively low because the output circuit includes the base-emitter circuit with its bias in the forward, or low resistance, direction.

The input impedance of the electron tube cathode follower is high and the output impedance is low, hence the cathode follower may be used as an impedance changer or stepdown transformer. The output signal, e_o, is in the series circuit between the cathode and grid and opposes the input signal, e_{in}, in the same circuit. Hence, the net voltage acting between the grid and cathode is $e_{in}-e_o$, and e_o must always be less than e_{in} if the grid-to-cathode voltage is to be a finite value. Thus the voltage gain of the cathode follower is ALWAYS less than 1.

Similarly, the output signal in the common-collector transistor opposes the input signal, and the net signal voltage acting between the base and emitter is $e_{in}-e_o$. For example, if the input signal causes the voltage occurring between the base and ground to swing 10 mv in a positive direction and the output voltage between emitter and ground to swing 4 mv in the same direction, the net voltage between base and emitter will be 10-4 or 6 mv. The output voltage, e_o, must always be less than e_{in} if the base-to-emitter voltage is to be a finite value. Thus the voltage gain of the

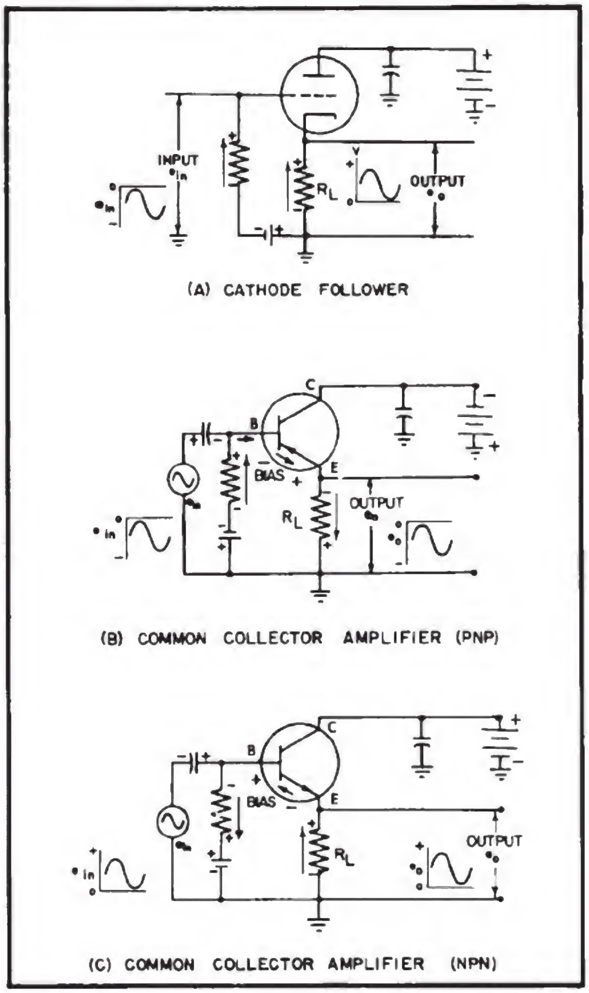

Figure 30-14 - Electron tube cathode follower and corresponding common-collector transistor amplifiers.

common-collector transistor, like that of the cathode follower, is always less than 1.

The common-collector transistor provides a relatively large current amplification and moderate power amplification, depending upon the magnitude of the output impedance.

The input impedance of the common-collector transistor amplifier is a function of the load impedance. This action is in contrast to that in the electron tube amplifier in which the input and output circuits are isolated and practically independent of each other. For example, the input impedance of a common-collector junction transistor amplifier may be on the order of 150 k ohms when the load impedance is 10 k ohms, but the input impedance may drop to 50 k ohms or less when the load impedance is reduced to 1000 ohms.

Table 30-1 lists typical ranges of values of resistances and gains for the various transistor configurations.

Quantity	Configurations		
	CB	CE	CC
Input impedance	30-150	500-1.5 k	20K-500 k
Output impedance	300 k-500k	30 k-50 k	50-1k
Voltage gain	500-1500	300-1000	less than one
Current gain	less than one	25-50	25-50
Power gain	20-30db	25-40db	10-20db

TABLE 30-1
Resistance and Gains.

Q10. Is the input impedance of a common-collector configuration a direct or inverse function of the load impedance?

30-10. Characteristic Curves

In the study of electron tubes it was found that the circuit conditions and operation could be predicted fairly accurately by use of a family of characteristic curves, and the method of developing these characteristic curves was discussed. Transistor circuits can also be analyzed by the use of characteristic curves. However, before using the curves in a circuit analysis, a brief discussion of their development may prove to be advantageous.

For electron tubes the familiar family of curves is the plate voltage (E_B)—plate current (I_B) curves in which the grid voltage (E_C) was the parameter or quantity held constant.

The usual curves used for transistors are the collector current (I_C)—collector voltage (V_C) curves in which the parameter is the emitter current (I_E) or the base current (I_B), depending on whether the common-base or common-emitter configuration is used in plotting the curves.

The method of plotting a family of characteristic curves will now be discussed. A test circuit is constructed in such a manner that the input current (I_E for a common-base configuration and I_B for a common-emitter configuration), and collector voltage are adjustable. Meters are connected into the circuit so that the currents and voltages can be monitored. Figure 30-15 shows such a test circuit.

Base resistor R_1 is adjusted so that 0.1 ma (100 microamperes) base current flows. This current is indicated on the milliammeter connected in the base lead. The arm of resistor R_o

A9. Alpha. In the common base the input current is I_E and the output current is I_C.

A10. Direct. Input impedance decreases as load impedance decreases.

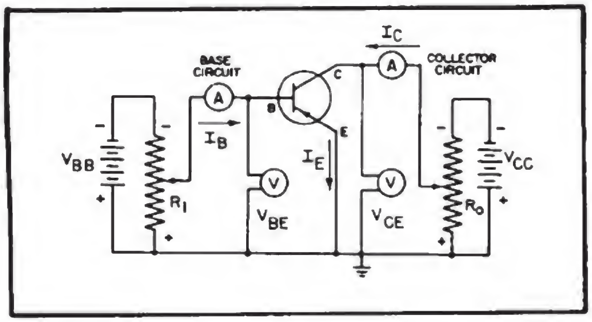

Figure 30-15 - Characteristic curves test circuit.

is moved to the end of the potentiometer connected to the junction of the emitter lead and the positive battery terminal, so that zero collector voltage is applied to the transistor. With the circuit adjusted in this manner the collector current is read on the milliammeter connected in the collector lead. I_C is read as zero. This is plotted as point one (P_1) on a graph of I_C versus V_{CE}, such as shown in Figure 30-16.

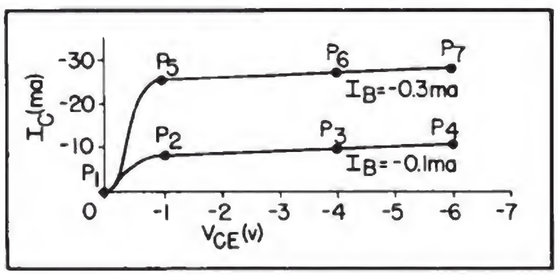

Figure 30-16 - Plot of I_C versus V_{CE} with I_B constant.

R_o is now adjusted until the voltmeter connected between the collector and the emitter indicates a V_{CE} of 1 volt (V_{CE}, I_B, and I_C are negative values on the graph because a PNP transistor is used). R_1 is again adjusted until a base current of 0.1 ma is flowing. The value of I_C is found to be about 8 ma and is plotted as P_2 on the graph (Figure 30-16). The procedure of increasing the collector voltage by adjusting R_o, holding the base current constant by adjusting R_1, and plotting the resultant collector current is continued for points 3 and 4; or as many points as are required. Finally, all the plotted points are connected with a line. The result is a graph which will allow the determination of the value of I_C for any value of V_{CE}, within the transistor rating, when the I_B is 0.1 ma.

The procedure is now repeated by setting V_{CE} back to zero (adjust R_o) and setting the base current at some new value such as 0.3 ma. Read the I_C and plot the value. Increasing the collector voltage in steps and holding the base current to 0.3 ma, points 5, 6, and 7 are plotted (Figure 30-16).

The above procedure is repeated until the I_C-V_C family of characteristic curves is developed. The same procedure is also used to develop a family of curves for a transistor used in the common-base configuration. The only variation being that the emitter current is now the value held constant while the collector voltage is varied.

Figure 30-17 is a comparison between a typical family of E_B-I_B characteristic curves for a pentode electron tube and a typical family of I_C-V_{CE} characteristic curves for a junction transistor. Many similarities between the two devices can be noted from the curves. In both cases current flow through the device increases rapidly at first and then levels off as collector or plate voltage is increased from zero. In both curves, for a given value of grid voltage or base current, the output current changes very little for large changes of supply voltage. The major current flow through both devices (at very low values of supply voltage) is a function of the supply voltage and not the control electrode.

30-11. Circuit Analysis

There are many instances in electronics when it is advantageous to be able to estimate or predict the action of a circuit from a relatively few known facts. Circuit analysis is greatly simplified by use of a family of characteristic curves and a LOAD LINE. Once a load line is established on the characteristic curves, it may be used to determine the operating point, biasing requirements, current gain, voltage gain, power gain, etc.

The procedure for establishing a load line is carried out in much the same manner as for electron tubes. The analysis of a circuit can best be conducted by an example problem.

Example. It is desired to analyze a transistor audio amplifier in which the following facts are known:

1. The collector supply voltage, V_{CC} is 15 V.
2. The load resistance, R_L, is 1875 ohms.
3. The transistor input resistance is 100 ohms.
4. The base resistance, R_B, is 29.9 kohms.
5. Base supply voltage, V_{BB}, is 3 V.

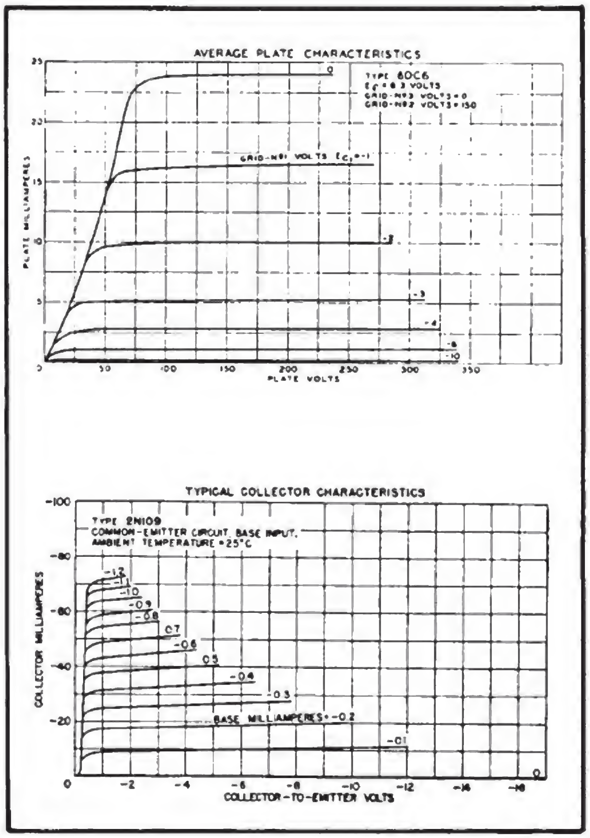

Figure 30-17 - Electron tube and transistor characteristic curves.

6. The transistor (Q_1) is a 2N464 medium gain, PNP, junction transistor.

Figure 30-18 illustrates the circuit to be analyzed.

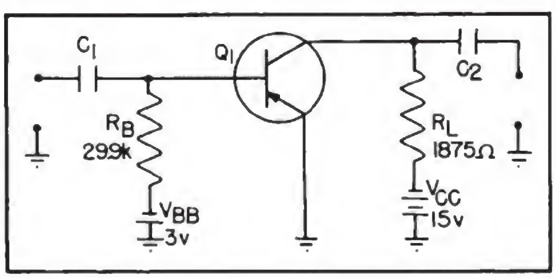

Figure 30- 18 - Audio voltage amplifier.

The collector characteristic curves for the 2N464 transistor, used in the audio amplifier, are shown in Figure 30-19. Figure 30-19 also shows the load line for the audio amplifier.

The procedure for establishing a load line for transistors is the same as for electron tubes. Two points are needed and since maximum collector voltage and maximum collector current are the most easily determined points they will be used. If no current flowed through R_L there would be no voltage drop across the load to subtract from the supply voltage. Therefore, the collector voltage, V_C, would equal the collector supply voltage. Since V_{CC} equals 15 V, and is a negative potential because of the PNP transistor, one end of the load line will be on the voltage axis at -15 V (point Z, Figure 30-19). Making the transistor appear as a short between the collector-emitter terminals will permit maximum current (limited only by R_L) to flow. The value of the collector current (I_C) can be determined by use of Ohm's law as follows:

$$I = \frac{E}{R}$$

$$I_C = \frac{V_{CC}}{R_L}$$

$$I_C = \frac{-15}{1875}$$

$$I_C = -8 \text{ ma}$$

The other end of the load line will be 8 ma (point Y, Figure 30-19). A straight line is drawn between points Y and Z to complete the load line.

The operating point occurs where the desired base current crosses the load line. Therefore, to find the operating point, the base current must be determined. The resistance of the input (base-emitter) circuit consists of R_B plus the input resistance of the transistor. The resistance is:

$$29,900 + 100 = 30 \text{ k ohms}$$

The base current can be found by employing Ohm's law as follows:

$$I = \frac{E}{R}$$

$$I_B = \frac{V_{BB}}{R}$$

$$I_B = \frac{3}{30,000}$$

$$I_B = 100 \text{ microamperes}$$

The 100 microampere base current line crosses the load line at point Q (Figure 30-19). Therefore, the operating point is established. Although the construction of the load line was a simple procedure, many new facts may now be determined about the operation of the circuit.

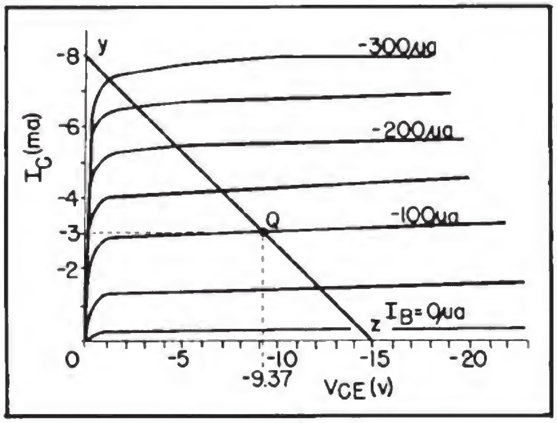

Figure 30-19 - 2N464 I_C-V_C curves and audio amplifier load line.

The input junction bias voltage, base-emitter, (V_{BE}), can be determined from the base current and the input junction resistance at the operating point.

$$V_{BE} = I_B \times r_i$$

$$V_{BE} = 1 \times 10^{-4} \times 1 \times 10^2$$

$$V_{BE} = 10 \text{ mv}$$

Thus, at the operating point or QUIESCENT POINT 10 mv of forward bias is applied to the transistor. Examination of Figure 30-19 will show that 3 ma of collector current is flowing under quiescent or no signal conditions.

Knowing the I_C at the operating point will permit determination of the collector voltage, since V_C equals V_{CC} minus the voltage drop across the load resistance. The load voltage drop at the operating point is:

$$E_{R_L} = I_C \times R_L$$

$$E_{R_L} = 3 \times 10^{-3} \times 1875$$

$$E_{R_L} = 5.625 \text{ volts}$$

Thus, the collector voltage at the operating point is:

$$V_C = V_{CC} - E_{R_L}$$

$$V_C = 15 - 5.63$$

$$V_C = 9.37 \text{ volts}$$

The emitter current is equal to the collector current plus the base current. Therefore, I_E equals:

$$I_E = I_C + I_B$$

$$I_E = 3 \text{ ma} + 0.1 \text{ ma}$$

$$I_E = 3.1 \text{ ma}$$

The use of the load line has simplified the analysis of the amplifier under quiescent (no signal) conditions.

Before applying a signal and analyzing the circuit, a brief summary of the known facts will be given. At the quiescent point (Q point) a forward bias of 10 mv causes an I_C of 3 ma to flow. The I_C flowing through R_L causes a voltage drop which, when subtracted from the supply voltage, established the V_{CE} at -9.37 volts. 100% of the current in the circuit flows through the emitter, 92% to 99% through the collector, and 1% to 8% through the base. In the audio amplifier of the example problem, the I_C equals 3 ma and the I_E equals 3.1 ma. Thus, the percentage of emitter current flowing in the collector is:

$$\% = \frac{I_C}{I_E} \times 100$$

$$\% = \frac{3}{3.1} \times 100$$

$$\% = 0.967 \times 100 \text{ or } 96.7\%$$

The base current is, therefore, 3.3% of the I_E.

The maximum peak input signal (current) that can be applied to the circuit is the value between the I_B=0 point and the I_B at the Q point or 100 microamperes. Because Q_1 is a PNP transistor, a positive going input voltage signal will oppose forward bias. If the input voltage causes a 200 microampere peak-to-peak current to flow in the input circuit, the base current will decrease to zero during the positive peak of the input voltage. Figure 30-19 shows that when I_B equals zero (on the load line) the I_C is reduced to approximately 0.5 ma (the reason that I_C is not zero will be explained in section 30-12). On the negative half cycle of the input voltage signal, when it aids the forward bias, the base current will equal 200 microamperes. The intersection of the 200 microampere base current and the load line occurs with 5.5 ma of collector current flowing. On the positive half cycle of the input voltage signal, the collector current decreased from 3 ma to 0.5 ma, or a swing of 2.5 ma. On the negative half cycle of the input signal the collector current increased from the 3 ma (Q) point to 5.5 ma, or a swing of 2.5 ma. Since the output current swings the same amount on each alternation, no distortion

is present. Thus, the load line can be used to estimate the maximum input signal that can be applied without producing distortion.

It was previously stated that alpha and beta were the theoretical maximum current gains of two types of transistor configurations. Alpha is the current gain of the common-base configuration and beta is the current gain of the common-emitter configuration. The alpha or beta of a transistor can be determined from the characteristic curves in the following manner:

The audio amplifier in the example problem employs the common-emitter configuration; therefore, its current gain will be beta. To be reasonably accurate, the beta must be determined by small current changes AROUND THE Q POINT. Beta has been defined as the ratio of the change in I_C to a corresponding change in I_B with the V_C being held constant. To determine beta, using Figure 30-19, extend a straight line through the operating point (point Q) and -9.37 V collector volts. Two convenient points for the change in base current are where the vertical line crosses the 50 microamp and the 150 microamp base current lines. This gives a change in I_B of 150-50 or 100 microamps. The I_C flowing when I_B is 50 microamps is 1.5 ma. When I_B is 150 microamps, I_C equals approximately 4.5 ma. This gives a change in I_C of 4.5 - 1.5 or 3 ma. Beta can now be determined using the equation:

$$B = \frac{\Delta I_C}{\Delta I_B}\Big|_{V_{CE}\ \text{constant}} \qquad (30\text{-}3)$$

$$B = \frac{3\ \text{ma}}{0.1\ \text{ma}}$$

$$B = 30$$

This indicates that the maximum possible current gain of this transistor (in the CE configuration) will be 30. In a practical circuit, with a load connected, the current gain will be something less than 30.

The current gain (A_i) is the peak-to-peak output current divided by the peak-to-peak input current. This can be expressed mathematically as:

$$A_i = \frac{\text{Output current}}{\text{Input current}} \qquad (30\text{-}6)$$

where: A_i = the current gain of the stage

output current = the peak-to-peak swing of I_C

input current = the peak-to-peak swing of input current

The peak-to-peak input current of the circuit in Figure 30-20 is 200 microamperes. Determining the value of collector current for zero base current and 200 microamperes base current (where the load line crosses these points) the output current is 5.5-0.5 ma or approximately 5 ma peak-to-peak. By equation (30-7), the current gain of the circuit is:

$$A_i = \frac{\text{Output current}}{\text{Input current}} \qquad (30\text{-}6)$$

$$A_i = \frac{5\ \text{ma}}{0.2\ \text{ma}}$$

$$A_i = 25$$

Figure 30-20 shows the audio voltage amplifier with current paths, polarities, voltages, etc. inserted. Notice that capacitors C_3 and C_4 have

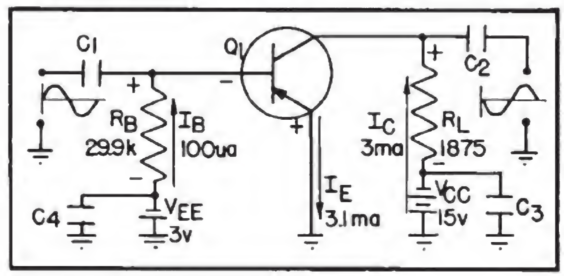

Figure 30-20 - Audio voltage amplifier.

been added as power supply bypasses. The signal waveforms are voltage waveforms.

Figure 30-21 shows the 2N464 characteristic curves with the load line, base current input waveform, collector voltage waveform, and collector current waveform for the audio amplifier.

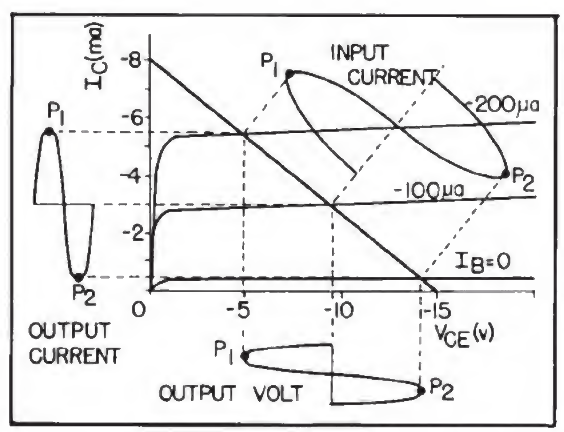

Figure 30-21 - 2N464 I_C-V_C curves with load line and waveforms.

Each waveform in Figure 30-21 has two points (P_1 and P_2) labeled to allow easier comparison. At point one the input and output currents are at their largest values and are in phase. However, the output voltage is at its lowest value at P_1. This is due, of course, to the large output current causing an increased E_{R_L} drop and subtracting more from V_{CC}. At point two both of the currents are at their lowest value, and the output voltage has increased to its highest value.

The voltage gain (A_v) and power gain (A_p) of the audio amplifier may be determined in the same manner as the current gain (A_i) by use of equations (30-7) and (30-8) shown below.

$$A_v = \frac{\text{Output voltage}}{\text{Input voltage}} \qquad (30\text{-}7)$$

where: A_v = voltage gain of the stage

output voltage = product of peak-to-peak collector current swing and R_L

input voltage = product of peak-to-peak input current swing and r_i

$$A_p = A_i \times A_v \qquad (30\text{-}8)$$

where: A_p = power gain of the stage
A_i = current gain of the stage
A_v = voltage gain of the stage

To determine the voltage gain of the audio amplifier (Figure 30-20), first determine the output voltage swing.

$$\text{Output voltage} = \text{output current} \times R_L$$

$$\text{Output voltage} = 5 \times 10^{-3} \times 1875$$

$$\text{Output voltage} = 9.375 \text{ volts}$$

Next determine the input voltage:

$$\text{Input voltage} = \text{input current} \times r_i$$

$$\text{Input voltage} = 2 \times 10^{-4} \times 1 \times 10^2$$

$$\text{Input voltage} = 20 \text{ mv}$$

Inserting these values into equation (30-8) yields:

$$A_v = \frac{9.375}{2 \times 10^{-2}}$$

$$A_v = 468.7$$

The power gain of the amplifier can be determined by application of equation (30-8) Insert known values:

$$A_p = A_i \times A_v \qquad (30\text{-}8)$$

$$A_p = 25 \times 468.7$$

$$A_p = 11,717$$

In some cases it may be desirable to have the power gain measured in db. This is easily accomplished in the following manner. Since the ratio of power out to power in is 11,717, then the power gain in db is:

$$A_p = 10 \log 11,717$$

$$A_p = 10 \times 4.0682$$

$$A_p = 40.68 \text{ db}$$

Q11. What circuit parameter is held constant when plotting a family of I_C-V_C curves for a common base configuration?

Q12. Assume the input junction resistance of Q_1 (Figure 30-20) increases from 100 ohms to 1 k ohms. It is desired to maintain the same operating point and base current. To what value must R_B be changed to accomplish this?

Q13. The same operating point is used (Figure 30-21), and the input current is reduced to a 100 microamp peak-to-peak signal. What would the approximate peak-to-peak output current be?

OPERATING LIMITS

Some of the limits imposed on an operating transistor are; maximum collector current, maximum collector voltage, cutoff current, saturation voltage, cutoff frequency, and maximum collector power dissipation. Even though a device has a maximum limit, for reasons of reliability, safety, and proper operation, it is not always operated at this maximum limit. This results in two areas of operating limits for transistors: (1) the maximum permissable operating region and (2) the linear operating region. As with electron tubes, the linear operating region is smaller than the permissable operating region.

30-12. Cutoff Current

CUTOFF CURRENT is the current that flows in the collector circuit when the input current is zero. It was previously shown that when the

emitter lead is opened and no current flows through the emitter-base (input junction), there will still be a small amount of current flowing in the collector circuit due to the reverse current I_{CBO}. I_{CBO} is the reverse current for a common-base configuration. If a common-emitter configuration is used, the reverse current will be termed I_{CEO}, or the current flowing in the collector circuit with the base lead open.

In any conductor, equal currents flowing in opposite directions will cancel and the resultant current in the conductor will be zero. Thus, in a common-emitter circuit an input signal that exactly opposes the forward bias will cause a zero base current. The result is the same as if the base lead had been opened. Even though the input current is zero, the output current will not necessarily be zero because of the reverse current. The area below the line marking the zero input current (either $I_E = 0$ or $I_B = 0$ depending on configuration used) is called the CUT-OFF REGION. The cut-off region is shown in Figure 30-22.

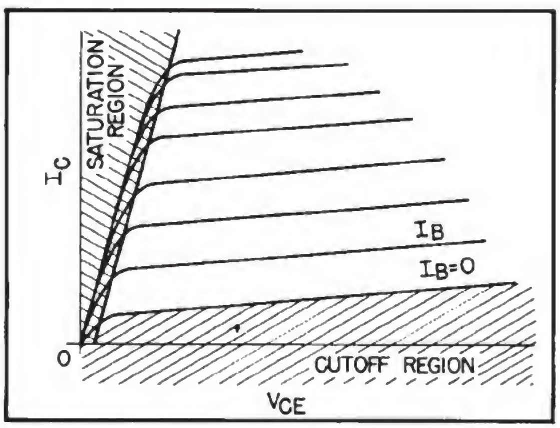

Figure 30-22 - Cut-off and saturation regions.

30-13. Voltage Saturation Region

In a common-emitter (CE) configuration, for a given value of base current, in the normal operating region, the collector current is affected very little by a change in collector voltage. However, at very low values of V_{CE}, the collector current changes drastically with a change in V_{CE}. The SATURATION REGION begins where the collector current begins to become a function of collector voltage rather than the input current (I_B or I_E). The saturation region is shown in Figure 30-22.

30-14. Maximum Collector Voltage

The MAXIMUM COLLECTOR VOLTAGE is the maximum dc potential that can be applied to the collector with safety and is usually specified by the manufacturer. Just beyond this value is the BREAKDOWN VOLTAGE (BV). The BV is the value of collector voltage at which the collector junction will break down and pass a high current, destroying the transistor. An example of the BV region is shown by the curves in Figure 30-23 which are the characteristic curves for the 2N1490 NPN silicon transistor. Notice that the BV must be specified for a certain value of I_B because the higher the I_B the lower is the breakdown voltage. For a base current of 5 ma, the I_B and I_C remain relatively constant from 1 volt to approximately 50 volts, V_{CE}. From 50 to 65 V_{CE}, both I_B and I_C begin increasing. Above 65 V_{CE} both currents increase very rapidly and the transistor breaks down. For higher values of I_B, the currents start to increase at a lower value of V_{CE}.

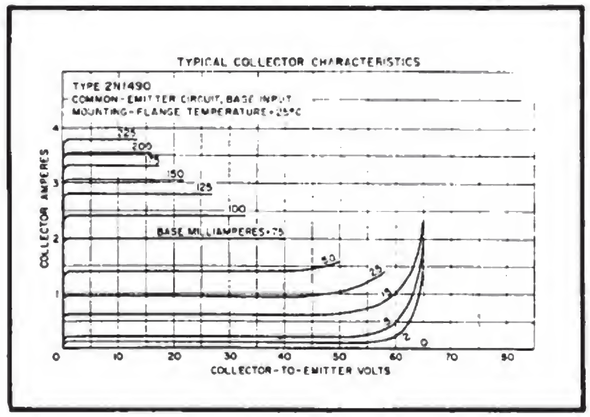

Figure 30-23 - 2N1490 curves showing BV

30-15. Maximum Collector Power Dissipation

MAXIMUM COLLECTOR DISSIPATION is the maximum power that may be safely dissipated by the collector. This operating limit is represented on the characteristic curves as a CONSTANT POWER DISSIPATION CURVE, MAXIMUM POWER DISSIPATION CURVE, P_{Dmax}, or TOTAL DISSIPATION.

P_{Dmax} represents the limit to the product of the dc quantities V_{CE} and I_C. In other words, if the manufacturer lists the maximum power dissipation of the transistor as 2 watts the product of collector voltage times collector current must not exceed 2 watts. Under ac conditions the instantaneous values of voltage or current may exceed this value as long as the average values do not.

A typical total dissipation curve is shown in Figure 30-24, on the characteristic curve of a 2N1067 NPN silicon transistor. The curve is graphed in the following manner:

Various values of V_{CE} are determined by dividing the manufacturers specified value of maximum power by various values of collector

A11. Emitter current, I_E.

A12. 29 k ohms. Since I_B is to remain 100 microamps and V_{BB} is to remain 3 volts, the total resistance of the junction plus R_B must remain 30 k ohms. Therefore, 30 k ohms minus 1 k ohm leaves R_B equal to 29 k ohms.

A13. 3 ma peak-to-peak.

---

current. For example:

In Figure 30-24 the maximum power dissipation is listed as 5 watts. Division of this value by 500 ma results:

$$V_{CE} = \frac{P_{Dmax}}{I_C} \qquad (30\text{-}9)$$

$$V_{CE} = \frac{5}{0.5}$$

$$V_{CE} = 10 \text{ volts}$$

A point is established at the intersection of the 500 ma I_C and the 10 volt V_{CE} lines. This procedure is continued for as many points as are necessary. When all the points are connected by a line, the power dissipation curve will be the result. A glance at this curve will give the maximum collector current that is permitted to flow for any value of collector voltage.

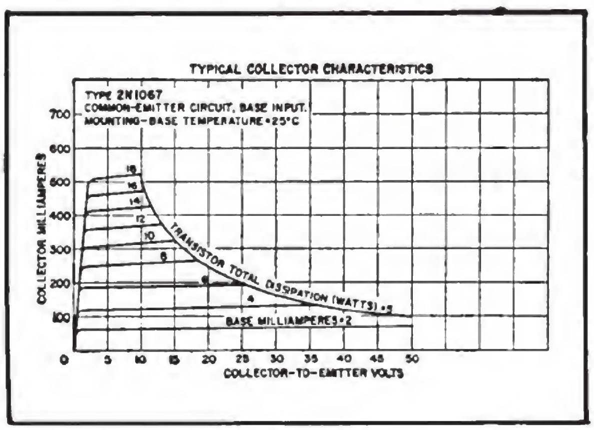

Figure 30-24 - Total dissipation curve for 2N1067 transistor.

The P_{Dmax} curve is also helpful in establishing the operating point and the load line because the load line and Q point must ALWAYS be to the left of the maximum power curve. The load line will sometimes fall tangent to the curve and sometimes well within the area, but NEVER to the right of the power curve.

30-16. Permissible Operating Area

From the previous sections the permissible operating area of a transistor is seen to be bounded by definite limits. This is the area within which the transistor may be safely operated. Figure 30-25 is a representative set of characteristic curves used to sum up the permissible operating limits. The heavy line, superimposed on the curves, is the boundry of the PERMISSIBLE OPERATING AREA.

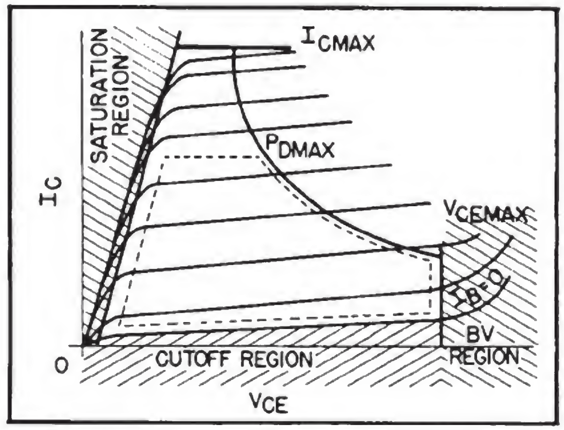

Figure 30-25 - Permissible operating area.

Current may vary from the maximum collector current to the collector cut-off. Voltage may vary from the voltage saturation limit to the breakdown voltage. The operating point must lie on or to the left of the power curve. The dotted line (Figure 30-25) indicates the LINEAR operating region which, as in electron tubes, is smaller than the permissible region.

Q14. Is the current flowing in the collector when the input current is zero due to majority or minority current carriers?

Q15. Would it be possible for a 2N1067 transistor (Figure 30-24) to operate under a no signal condition of 400 ma collector current and 25 volts collector to emitter?

CLASSES OF OPERATION

As in the case of electron tube amplifiers, transistor amplifiers are also operated class A, class B, class AB, or class C.

30-17. Class A Operation

Transistor amplifiers operated class A are biased in such a manner as to allow collector current flow during a complete 360 degrees of the input signal. Collector current also flows when no signal is present. The operating point is established, by proper biasing, so that the transistor will operate over the linear portion of its collector characteristics; thus, providing an output waveform which is an exact replica of the input waveform. Figure 30-26 shows the load line and operating or Q point of a 2N464 transistor operated class A. The input voltage waveform (dotted line) and the input current waveform (solid line) are out of phase because a PNP transistor is being used; therefore, a negative going input voltage will cause an increase in base current. The output current waveform has the same shape as the input current waveform. To insure class A operation the forward bias of the input junction must be equal to (or larger than) the maximum peak amplitude of the input signal voltage. Thus, the input junction will never be reverse biased.

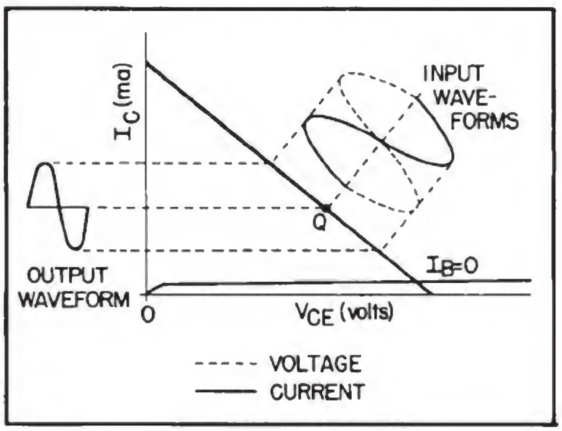

Figure 30-26 - 2N464 transistor, class A operation.

30-18. Class AB Operation

Transistor amplifiers for class AB operation are biased so that I_C is zero for a portion of one alternation of the input signal (collector current will flow for more than 180° but less than 360°). The forward bias voltage of the input junction is less than the peak value of the maximum input signal. Therefore, the input junction will be reverse biased for a portion of one alternation of the input signal. The junction will remain reverse biased for the amount of time that the positive value of signal voltage EXCEEDS the value of forward bias voltage. Figure 30-27 shows the characteristic curves of a 2N464 type transistor being operated class AB.

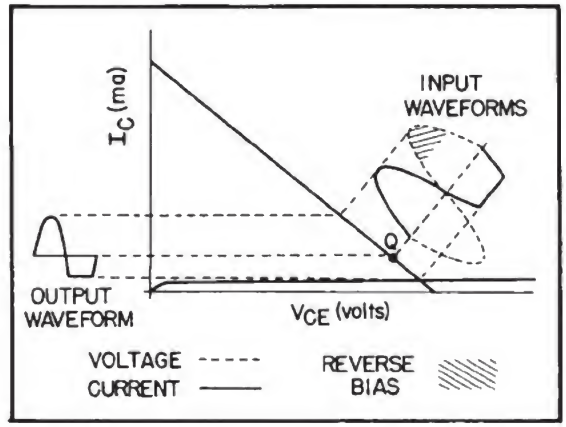

Figure 30-27 - 2N464 transistor, class AB operation.

Examination of the input waveforms indicate that the operating or Q point is such that a portion of the positive half cycle of the signal voltage (dotted line) will reverse bias the base-emitter junction. The signal current waveform (solid line) shows the base current as being equal to zero during the period of reverse bias on the B-E junction. The output current is very small during the time I_B is zero, because collector current during this period is due to reverse current flow I_{CEO}.

30-19. Class B Operation

Figure 30-28 shows a 2N464 transistor with the operating, Q, point at the intersection of the load line and zero base current, for class B operation.

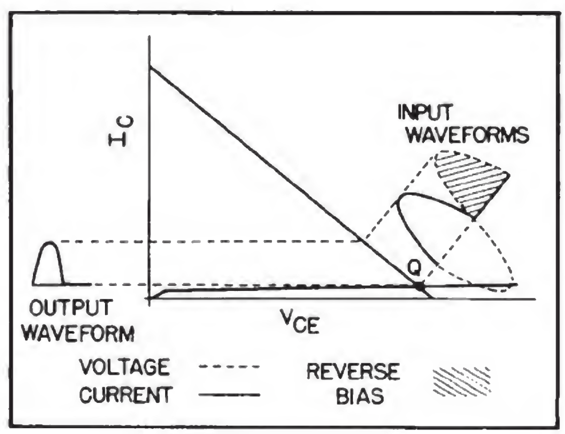

Figure 30-28 - 2N464 transistor, class B operation.

When a transistor amplifier is biased class B collector current will be cut-off during one half of the input signal. The operating or Q point is

A14. Minority because it is a reverse current.

A15. No. Collector dissipation would be 10 watts, and the Q point would lie to the right of the power curve.

---

established, by B-E junction bias, so that base current is zero under "no signal" conditions. When a signal is applied, one half cycle will forward bias the B-E junction and I_C will flow, while the other half cycle will reverse bias the B-E junction and I_C will be cut-off. Thus, for class B operation collector current will flow for 180 degrees of the input signal.

The positive half cycle of the input voltage reverse biases the PNP transistor B-E junction.

Class B operation has the advantages of very low collector current under no signal conditions and better power efficiency than class A or AB, however, a class B amplifier produces severe signal distortion.

30-20. Class C Operation

Class C amplifiers are biased so that collector current is zero for MORE than 180 degrees of the input signal. In other words, the transistor remains in a cut-off condition for all but a small portion of the input signal. To establish an operating point below cut-off in a transistor requires that the base-emitter junction be REVERSED BIASED. With the input junction reverse biased only the portion of the input signal that overcomes this bias will cause collector current to flow.

Figure 30-29 shows the characteristic curves and load line of a 2N464 transistor with the operating or Q point established below collector current cut-off. Notice that only the top portion of the input signal is negative enough to overcome the positive reverse bias on the input junction of the PNP transistor and cause current flow.

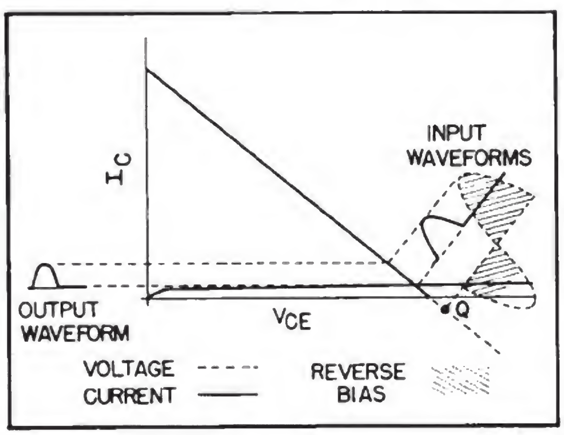

Figure 30-29 - 2N464 transistor, class C operation.

Q16. If a transistor is to be operated class A, what would be the minimum value of bias used with a 200 mv peak-to-peak input signal?

Q17. If an NPN transistor is operated class B, would the negative or positive half cycle of the input signal cause conduction? (CE Configuration)

Q18. What would be the polarity of V_{BB} for an NPN transistor (CE configuration) to produce class C operation?

ADDITIONAL BIAS ARRANGEMENTS

Proper biasing of a class A transistor amplifier, regardless of the circuit configuration used, consists of forward bias on the emitter base junction and reverse bias on the collector-base junction. In the previous bias arrangements two batteries were employed. However, it is possible to achieve proper biasing with only one battery.

30-21. Single Battery Bias, CB Configuration

In order to operate the common-base amplifier with a single battery, a voltage-divider network is required. The collector-base bias is achieved directly by the battery in the CB circuit. Since the transistor shown is a PNP type, reverse bias is achieved by making the collector negative with respect to the base. This is shown in Figure 30-30. Forward bias in the E-B circuit of the PNP transistor requires that the emitter be positive with respect to the base. This condition is achieved by the voltage divider network consisting of resistors R_3 and R_4. The

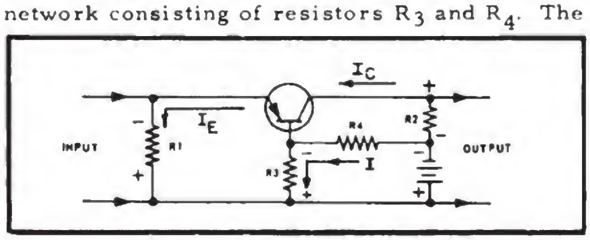

Figure 30-30 - Single battery bias, CB amplifier.

electron current (I) from the battery flows through the voltage divider in the direction

shown. This current flow causes a voltage drop across resistors R_3 and R_4 with the polarities as indicated. The voltage drop across R_3 places the emitter at a positive potential with respect to the base. To use this bias arrangement with an NPN transistor merely reverse the battery.

30-22. Single Battery Bias, CE Configuration

The common-emitter amplifier may also be biased with a single battery as shown in Figure 30-31. The single battery directly produces the required reverse bias voltage in the collector-base circuit. To understand the method by which the forward bias between the emitter and the base is produced by the single battery, a knowledge of the internal structure of the transistor is required. It has been stated previously that forward bias for the PNP transistor requires the base to be negative with respect to the emitter. In a PNP transistor, the collector is at the highest negative potential; the emitter is at the highest positive potential. Structurally the base is between the two, and therefore must assume a voltage between the two. Thus, the base must be less positive than the emitter or, in other words, negative with respect to the emitter. This condition satisfies the requirement of polarity necessary to produce a forward bias.

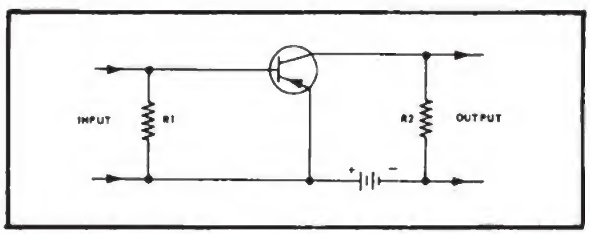

Figure 30-31 - Single battery bias, CE amplifier.

The magnitude of the voltage between the emitter and the base must be very small compared to that between the collector and base. Internally the two PN junctions act as a voltage divider. The PN junction between the collector and the base represents a high resistance and develops the larger voltage drop. The PN junction between emitter and base represents a low resistance and develops a low voltage as required for forward bias.

Another method of providing bias for a CE amplifier, using a single battery, is shown in Figure 30-32. In this case, R_1 and R_B form a voltage divider with R_B being much smaller than R_1. The forward bias is developed across the input junction and not across R_B. Since an NPN transistor is used, the emitter must be negative with respect to the base and the base positive with respect to the collector. The current paths and voltage polarities are shown in Figure 30-32.

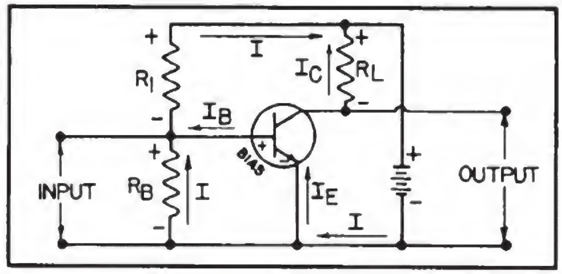

Figure 30-32 - Single battery bias, CE amplifier.

Q19. The circuit of Figure 30-31 depends on I_{CBO} to supply forward bias. Would this arrangement by satisfactory under varying temperature conditions? Why?

Q20. In the single battery bias arrangement, for the CB amplifier, (Figure 30-30) is the magnitude of the forward bias voltage the magnitude of the voltage drop across R_3?

BIAS STABILIZATION

In many ways the biasing of an electron tube presents less problems than biasing transistors. Negative grid biasing is used in most electron tube circuits. Because of this the selection of grid circuit resistances does not have to take the effect of grid current into consideration. The plate current of an electron tube relies on electrons generated by a heated cathode, the temperature of which is many degrees hotter than the ambient temperature. Therefore, a variation in the ambient temperature will have very little effect on the generation of electrons or plate current in an electron tube circuit.

Transistors, on the other hand, are very sensitive to ambient temperature variations. Reverse current, junction resistances, and collector current vary greatly with ambient temperature. Considerable care must be exercised in the selection of resistances in the input circuit and base lead due to the base current caused by forward bias.

774-639 O - 65 - 4

A16. 100 mv minimum forward bias to prevent the signal from reverse biasing the input junction.

A17. Positive half cycle.

A18. Negative to the base element and positive to the emitter element.

A19. No. I_{CBO} varies greatly with temperature and would thereby vary the bias.

A20. No. The forward bias is developed across the B-E junction resistance, which is much smaller than R3. Therefore, the forward bias voltage will be much smaller than the E_{R_3} drop.

30-23. Operating Point

The bias (operating point) for a transistor is determined by specifying the quiescent (no signal) values of collector voltage and emitter current. To have reliable operation of a transistor over a wide range of temperatures requires stable bias voltages and currents. In an NPN transistor, reverse bias collector current or saturation current (I_{CBO}) is the flow of holes (minority carriers) from the collector region toward the base region. The holes from the collector can accumulate in the base region if the resistance of the base region is high, or if high value external resistors are connected to the base.

As the holes increase in number, emitter electron flow into and through the base increases. This, in turn, increases the collector current. Increasing the collector current raises the temperature of the collector-base junction increasing the saturation current I_{CBO}. Figure 30-33 is a graph of the relationship between temperature and reverse current. It is possible for temperature and current to increase until severe distortion occurs, the transistor becomes inoperative, or the transistor destroys itself. This action is termed THERMAL RUNAWAY. To help prevent this condition, high-valued resistors should not be used in the base lead. The condition just described is the same for a PNP transistor, except that the parts the electrons and holes play are interchanged.

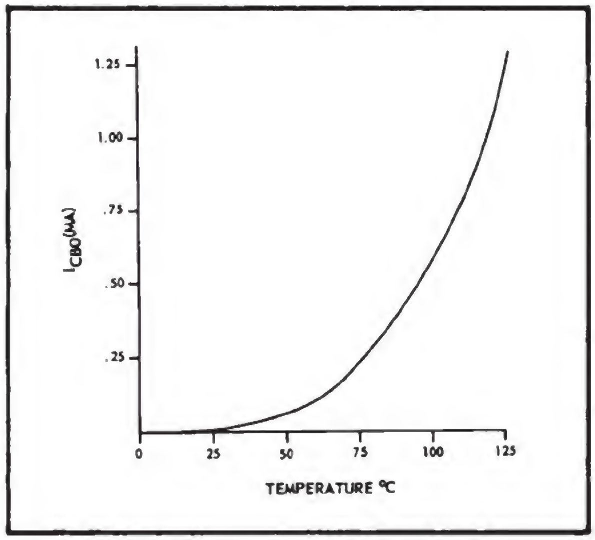

Figure 30-33 - Variation of I_{CBO} with temperature.

It was shown (Figure 30-33) that reverse current is negligible below 10 degrees centigrade. If only the reverse current caused the collector current variations with temperature, then the collector current should be constant at temperatures below 10° C. However, this is not the

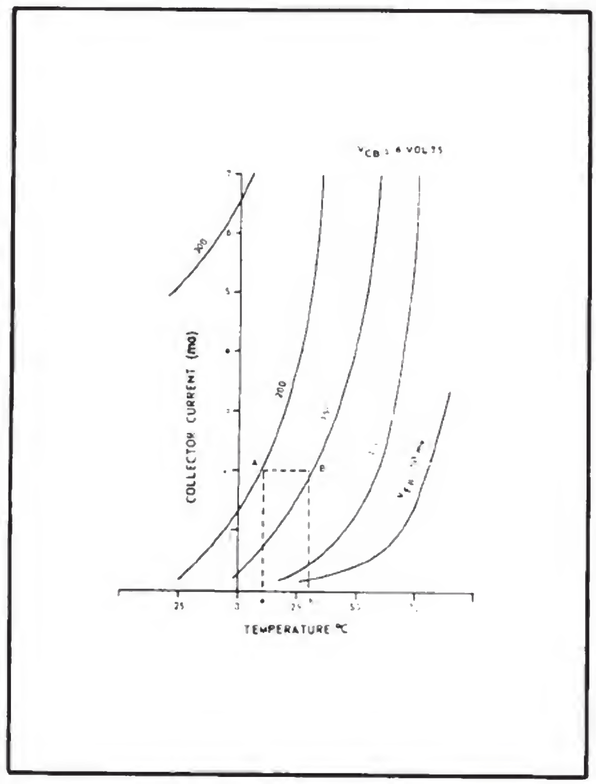

Figure 30-34 - Variation of I_C with temperature.

30-24. Emitter-Base Junction Resistance

The family of curves in Figure 30-34 indicates the variation of collector current with temperature. Each curve is plotted with a fixed collector-base voltage (V_{CB}) and a fixed emitter-base voltage (V_{EB}).

case because collector current does vary with temperature even when the reverse current is close to zero. This variation is caused by the decrease in emitter-base junction resistance when the temperature is increased. This means that the emitter-base junction resistance has a NEGATIVE TEMPERATURE COEFFICIENT of resistance. Junction resistance is an inverse function of temperature. One method of reducing the effect of the negative temperature coefficient of the junction is the insertion of a SWAMPING RESISTOR (large value resistor) in the emitter lead. The junction resistance is many times smaller than the emitter resistor (R_E), and in effect, any variation of the EB junction resistance becomes a small percentage of the TOTAL resistance in the emitter circuit. The name swamping resistor, for R_E, comes about because the external resistor is said to SWAMP or overcome the variation in the junction resistance.

Another method of reducing the effect of the negative temperature coefficient or thermal runaway is to reduce the EB bias as the temperature increases. This method is illustrated in Figure 30-34. In order to maintain the collector current at 2 ma, while the transistor temperature varies from 10° (point A) to 30° C (point B), the forward bias must be reduced from 200 mv (point A) to 150 mv (point B). The temperature difference is approximately 20° C; the voltage difference is 50 mv. The variation in forward bias per degree centigrade is calculated in the following manner:

$$\frac{\text{Difference in forward bias}}{\text{Difference in temperature}} \qquad (30\text{-}10)$$

$$= \frac{50 \text{ mv}}{20^\circ \text{ C}}$$

$$= 2.5 \text{ mv}/^\circ \text{C}$$

All maximum ratings for transistors are usually given at 25°C. If the transistor is to be operated at a higher temperature the maximum ratings will decrease. In order to determine the value of bias that will establish an operating point within the operating limits at the higher temperature a DEGRADATION FACTOR (correction factor) is used. This degradation factor is stated in mv/°C of temperature raise and is determined as shown above.

Q21. A transistor is being operated with a V_{EB} of 150 mv and an I_C of 4 ma. The ambient temperature is approximately 45°C. If the temperature decreases will the forward bias have to be increased or decreased to maintain I_C at 4 ma?

Another way of expressing the change of transistor characteristics with temperature variations is by use of a family of characteristic curves. Figure 30-35 shows a family of curves for a 2N464 transistor with a load line and operating point established for normal operating temperature. 100 microamps of base current causes 3 ma of collector current and a collector voltage of 9.37 volts. The effect of a temperature increase is to cause the spacing between the base current curves to spread out. In other words, maintaining the same base current flow under increased temperature conditions will cause more collector current to flow. A larger collector current flowing through the load resistance will cause a larger E_{RL} drop which, in turn, will lower the collector voltage. The current and voltage will change due to the shift of the operating point with a temperature increase.

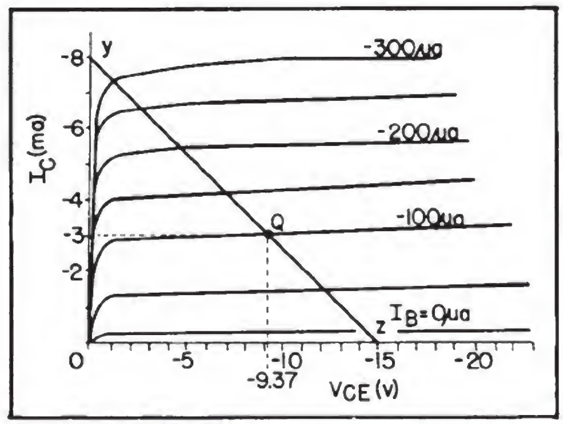

Figure 30-35 - Operating point under normal temperature conditions.

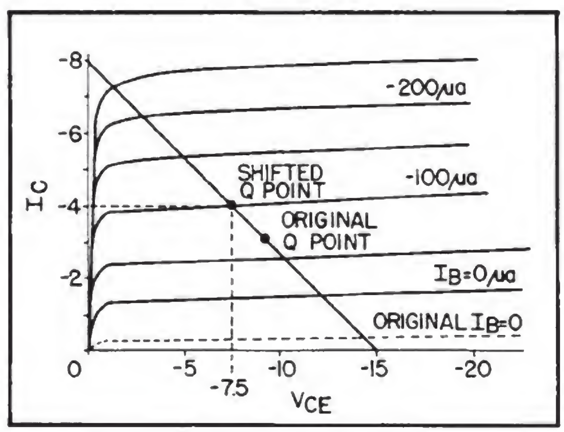

Figure 30-36 - Operating point under increased temperature conditions.

Figure 30-36 shows the 2N464 transistor operating under increased temperature conditions.

A21. Increase the EB bias an appropriate amount to restore the proper operating conditions.

Notice the increased spacing between the base current curves and the shift of the operating point.

The collector current has increased to approximately 4 ma while the collector voltage has decreased to approximately 7.5 volts.

Figure 30-37 shows the transistor operating under decreased temperature conditions. The spacing between the base current curves has decreased, the operating point has shifted lower on the load line, less collector current is flowing, and the collector voltage has increased.

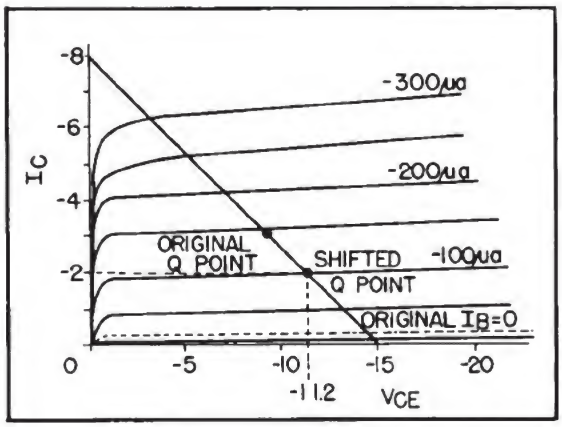

Figure 30-37 - Operating point under decreased temperature conditions.

Q22. Give one method of compensating for a shifted Q point due to a temperature increase.

30-25. Resistor Stabilization

In the transistor amplifier illustrated in Figure 30-38, resistor R_B provides fixed bias for class A operation, and resistor R_E provides circuit stabilization to prevent temperature changes from altering the transistor characteristics. Resistor R_E also provides protection to the transistor by limiting the magnitude of the collector current to the maximum safe value for the highest temperature to be encountered during operation.

If the temperature increases, the increase in collector current through R_E will lower the base-emitter voltage. This action will tend to prevent further increase in collector current, and therefore, minimize the shift of the operating point. Conversely, a decrease in temperature will lower the collector current through R_E and the voltage drop across R_E. This action will increase the base-emitter voltage and will tend to prevent further reduction in collector current so that there will be less shift of the operating point.

The value of R_E is approximately equal to the value of R_L for maximum protection. However, this would result in a serious reduction of power output efficiency since R_E would dissipate as much power as R_L. It is believed that most transistor audio power amplifier circuits will be satisfactorily stabilized if the value of R_E is not more than 10 percent of the load resistance (providing the R_L was designed for maximum power output).

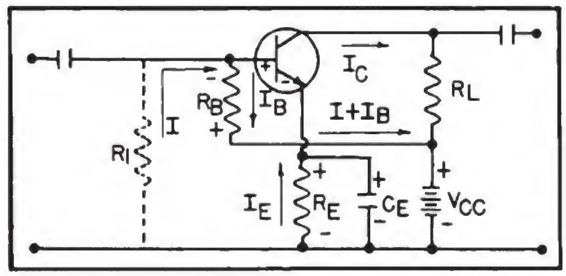

Figure 30-38 - Resistor bias stabilization.

If resistor R_1 is added to the circuit of Figure 30-38, another return path for battery current will be provided. This path is around R_E. The effect of R_1 is to provide a relatively fixed base bias voltage that is independent of current change due to temperature effects. The total current in R_B is the sum of the current through R_1 and the base current I_B.

Increases in temperature tend to increase I_B and I_C and in the absence of R_1 might damage the transistor.

With R_1 in the circuit, temperature induced changes produce less voltage change across R_B due to the voltage divider action. At the same time the average dc voltage across R_E will oppose the voltage across R_1 (like negative feedback) to compensate for the temperature effect. For example, an increase in temperature will cause an increase in collector current and voltage drop across R_E. The voltage across R_1 is equal to the sum of the voltage across R_E and the bias voltage across the base-emitter junction. The voltage across R_E increases directly with collector current. However, as mentioned before, the voltage across R_1 is approximately constant because of the voltage divider action of R_1 and R_B. Thus, an increase in voltage across R_E is accompanied by a decrease in forward bias voltage and current, and the rise in I_C is limited. On the other hand more power will be wasted at the input, but there will be less change in input impedance and less change

in input circuit loading. The net result is that this method of bias sacrifices a little power gain in return for better stabilization.

In order to prevent degeneration from occurring across the stabilizing resistor, R_E (Figure 30-38), in the emitter circuit, a bypass capacitor, C_E, is connected in parallel with R_E. The resulting action is similar to that occurring in the bypass capacitor in parallel with the cathode resistor of a cathode biased electron tube amplifier. In order to bypass the ac component around the resistor without developing a voltage at the signal frequency across the resistor, the X_C ohms of the bypass capacitor should be low with respect to the resistance of the resistor. For most audio transistor amplifiers the bypass capacitor does not need to be larger than about 50 microfarad.

Q23. If capacitor C_E in Figure 30-38 were to become shorted, would the forward bias increase or decrease? Why?

30-26. Thermistor Stabilization Circuits

It has been established that the bias current of the transistor is temperature sensitive. Specifically, emitter current increases with an increase in temperature. Emitter current stabilization can be achieved by use of external circuits using temperature sensitive elements. There are several temperature sensitive electrical elements. One such element is the THERMISTOR (contraction of the words thermal and resistor). The thermistor, as used in this chapter, has a negative temperature coefficient of resistance; that is, its resistance value decreases with an increase in temperature.

The circuit shown in Figure 30-39 employs a thermistor to vary the emitter voltage with temperature to minimize temperature variations in emitter current. This circuit contains two voltage dividers, the first consisting of resistors R_2 and R_1, and the second consisting of resistor R_E, and thermistor RT_1. The first voltage divider permits the application of a portion of battery V_{CC} voltage to the base terminal and ground (common return). The base terminal voltage is developed across resistor R_1 and is in the forward bias direction. The second voltage divider applies a portion of battery V_{CC} voltage to the emitter terminal. The emitter terminal voltage is developed across resistor R_E and is in the reverse bias direction. The forward bias voltage applied to the base terminal is larger than the reverse bias applied to the emitter terminal, so that the resultant base-emitter bias is always in the forward direction.

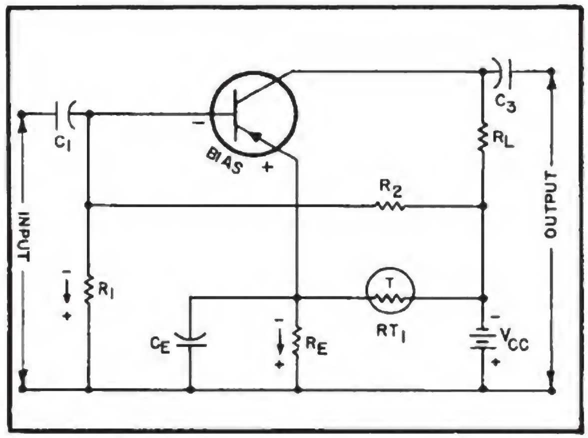

Figure 30-39 - Thermistor control of emitter bias voltage.

With an increase in temperature, the collector current would normally increase if the transistor were not stabilized. The increase in collector current can be prevented by reducing the forward bias. This is accomplished by the voltage divider consisting of resistor R_E and thermistor RT_1. As the temperature increases, the resistance of thermistor RT_1 is decreased, causing more current to flow through the voltage divider. The increased current raises the negative potential at the emitter connection of resistor R_E. This action increases the reverse bias applied to the emitter and decreases the net emitter-base forward bias. As a result, the collector current is reduced. Similarly, decreasing the temperature would cause the reverse actions, and prevent the decrease of collector current.

Capacitor C_1 blocks the dc voltage of the previous stage and couples the ac signal into the base-emitter circuit. Capacitor C_E bypasses the ac signal around resistor R_E. Resistor R_L is the collector load resistor and develops the output signal. Capacitor C_3 blocks the dc collector voltage from, and couples the ac signal to, the following stage.

The circuit shown in Figure 30-40 employs a thermistor to vary the base voltage with temperature to minimize temperature variations in emitter current. This circuit contains a voltage divider consisting of resistor R_1 and thermistor RT_1. The voltage divider applies a portion of battery (V_{CC}) voltage to the base-emitter circuit. Electron current flow through the voltage divider is in the direction of the arrow. This current produces a voltage of the polarity indicated across thermistor RT_1. This circuit produces a forward bias.

If the temperature of the transistor rises, the emitter current would tend to rise. However, the resistance of thermistor RT_1 decreases with

A22. Reduce the EB bias an appropriate amount to restore the proper operating conditions.

A23. Increase. The bias would now be approximately equal to the voltage drop across R_1. The transistor would most likely be destroyed due to thermal runaway.

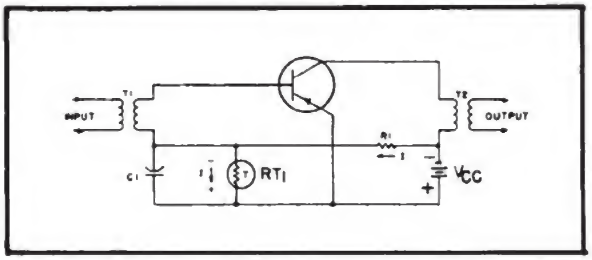

Figure 30-40 - Thermistor control of base bias voltage.

increase in temperature, causing more current to flow through the voltage divider. Increased current through the voltage divider causes a larger portion of battery V_{CC} voltage to be dropped across resistor R_1. The available voltage for forward bias developed across the thermistor RT_1, is reduced, thereby reducing the emitter current.

Transformer T_1 couples the ac signal into the base-emitter circuit. Capacitor C_1 bypasses the ac signal around thermistor RT_1. Transformer T_2 primary acts as the collector load and develops the output signal which is coupled to transformer T_2 secondary.

The ability of the thermistor to limit the variation of collector current with temperature is represented in Figure 30-41. One curve shows the variation of collector current for a transistor circuit that is not stabilized. Another curve shows the variation of collector current for a transistor circuit that is resistor stabilized. The last curve shows the variation of collector current for a transistor circuit that is thermistor-stabilized. There is improvement in the stability of the transistor circuit employing the thermistor. However, thermistor stabilization achieves ideal current at only three points (A, B, and C, Figure 30-41) because the thermistor resistance variation is not equal to (does not track) the variation in emitter-base junction resistance.

Q24. If power output efficiency is a main consideration of a circuit, would it be better to stabilize the circuit by thermistor control of the emitter bias voltage or the base bias voltage?

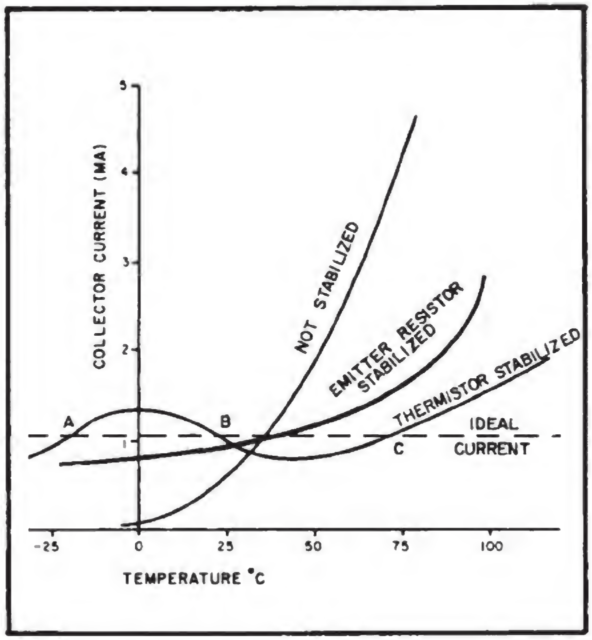

Figure 30-41 - I_C versus temperature for non-stabilized and thermistor stabilized circuits.

Q25. If a transistor circuit is required to operate with an ambient temperature of about 75° C would resistor or thermistor stabilization be used?

30-27. Junction Diode Stabilization

Variation of collector current with temperature is caused by the variation with temperature of the emitter base junction resistance and the saturation (reverse bias) current. Variation, with temperature, of the resistance and the reverse bias current of a PN junction occurs whether the PN junction is part of a transistor or part of a junction diode. Thus, the junction diode, like the thermistor, can be used in bias stabilizing circuits. The main advantage of using the junction diode as a temperature-sensitive element is that it can be made of the same material as that of the transistor to be stabilized. The temperature coefficient of resistance of the diode and that of the transistor, of the same material, are the same. This condition permits a more constant collector current over a wide range of temperatures because of better tracking with the EB junction resistance.

Junction diodes have a negative temperature coefficient of resistance whether they are forward or reverse biased (as long as the breakdown region is avoided).

30-28. Forward Biased Single Diode Stabilization

The circuit shown in Figure 30-42 employs forward-biased junction diode CR_1 as a temperature sensitive element to compensate for variations of emitter-base junction resistance. Consider the voltage divider consisting of resistor R_1 and junction diode CR_1. The current (I) through the voltage divider flows in the direction shown, and develops a voltage across diode CR_1 with polarity as indicated. This voltage is a forward bias. With increased temperature the collector current would tend to increase. However, increased temperature decreases the resistance of diode CR_1, causing more current to flow through the voltage divider.

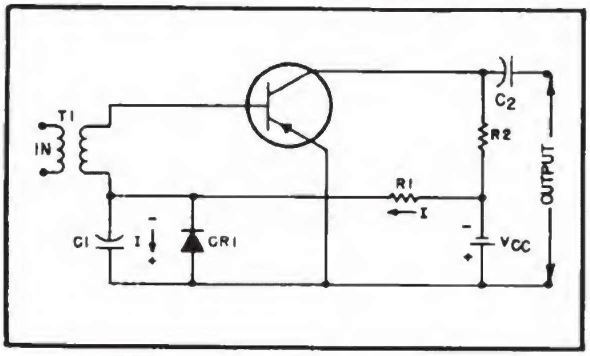

Figure 30-42 - Single forward-biased diode stabilization.

As a result there is an increased voltage drop across resistor R_1. The voltage drop across diode CR_1 is correspondingly decreased, thereby reducing the forward bias and hence, the collector current.

The ac signal is coupled into the transistor amplifier by transformer T_1. Capacitor C_1 bypasses the ac signal around diode CR_1. Collector load resistor R_2 develops the output signal. Capacitor C_2 blocks the dc voltage from, and couples the ac signal to, the following stage.

The effectiveness of this circuit to stabilize collector current with temperature is indicated by curve BB of Figure 30-43. Compare this curve with curve AA, for which the transistor is not stabilized and with the curve for ideal current. Curve BB shows that a marked improvement in the collector current stability occurs for temperatures below 50° C. This indicates that the variations with temperature of the junction diode resistance tracks and compensates for the variations of emitter-base junction resistance. The sharp increase in collector current (curve BB) at temperatures above 50° C indicates that junction diode CR_1 does not compensate for the increase in saturation current (I_{CBO}). This condition might have been anticipated since the saturation current (collector-base reverse bias current) flows out of the base, through T_1 secondary, through diode CR_1, battery V_{CC}, and back to the collector. Because the saturation current, at temperatures below 50° C, is a small percentage of the total current through diode CR_1 (forward biased and low in resistance) it causes no appreciable voltage drop across diode CR_1.

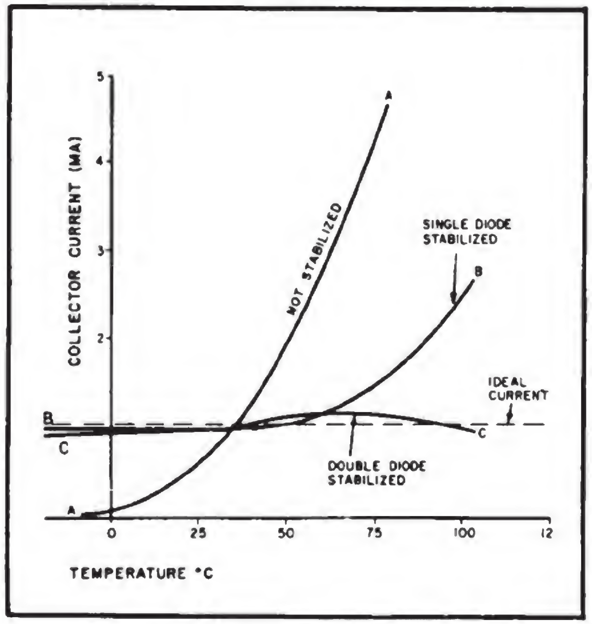

Figure 30-43 - I_C versus temperature for non-stabilized, single diode and double diode stabilization.

30-29. Double Diode Stabilization

The circuit shown in Figure 30-44 employs two junction diodes as temperature sensitive elements. One junction diode compensates for the temperature variations of emitter-base junction resistance; the other compensates for the temperature variations of saturation current. The circuit is similar to that shown in Figure 30-42. Resistor R_3 and junction diode CR_2 (reverse biased) have been added (Figure 30-44). Resistor R_1 and junction diode CR_1 (forward

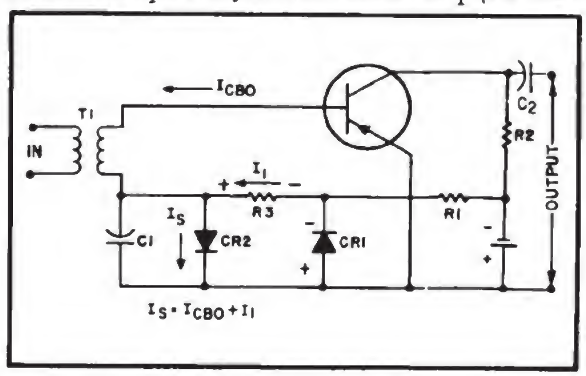

Figure 30-44 - Double diode stabilization.

A24. Thermistor control of the base bias voltage is better because power is wasted in the emitter resistance.

A25. Thermistor stabilization, because it is closer to the ideal current at higher temperatures.

---

biased) compensate for the change in emitter-base junction resistance at temperatures below 50^{o} C. (See curve BB, Figure 30-43.)

Reverse-biased junction diode CR_2 can be considered an open circuit at low temperatures. At room temperature, diode CR_2 reverse bias (saturation) current I_s flows through junction diode CR_2 in the direction indicated. Diode CR_2 is selected so that its rated saturation current is larger than the rated I_{CBO} of the transistor. Diode reverse-bias current I_s consists of transistor reverse bias current I_{CBO} and a component of current (I_1) drawn from the battery. The voltage polarity developed by current I_1 across resistor R_3 is indicated. Note that the emitter-base bias voltage is the sum of the opposing voltages across resistor R_3 and junction diode CR_1, assuming negligible resistance in transformer T_1 secondary. As the temperature increases, I_{CBO}, I_s, and I_1 increase. The resultant reverse-bias voltage developed across resistor R_3 by current I_1 increases. The total forward bias (voltage across diode) CR_1 and resistor R_3 decreases with increasing temperature to stabilize the collector current.

The functions of transformer T_1, capacitors C_1 and C_2, and resistor R_2, are the same as those for the corresponding elements of Figure 30-42.

The effectiveness of this circuit to stabilize collector current at high and low temperatures is indicated by curve CC of Figure 30-43.

30-30. Breakdown Diode Temperature Compensation

It was stated that the reverse-biased PN junction diode has a negative temperature coefficient of resistance. This statement is true if the reverse-bias voltage does not equal or exceed the breakdown voltage. The breakdown diode (AVALANCHE DIODE) has a positive temperature coefficient of resistance several times larger than the negative temperature coefficient of resistance of the forward- or reverse-biased PN junction diode.

Voltage regulators using breakdown diodes were discussed in the previous chapter. The discussion of the operation of a breakdown diode voltage regulator applies only if the temperature of the breakdown diode does not vary under operating conditions. One method of compensating for the increasing or decreasing breakdown diode resistance with increasing or decreasing temperature, respectively, is to place elements of negative temperature-coefficient in series with the breakdown diode. Figure 30-45 shows a circuit with two forward-biased diodes (CR_1 and CR_2) in series with the breakdown diode (CR_3). The total resistance of the three diodes in series remains constant over a wide range of temperatures. The complete result is a constant voltage output although temperature, input voltage (E_{in}), and load-current drain may vary.

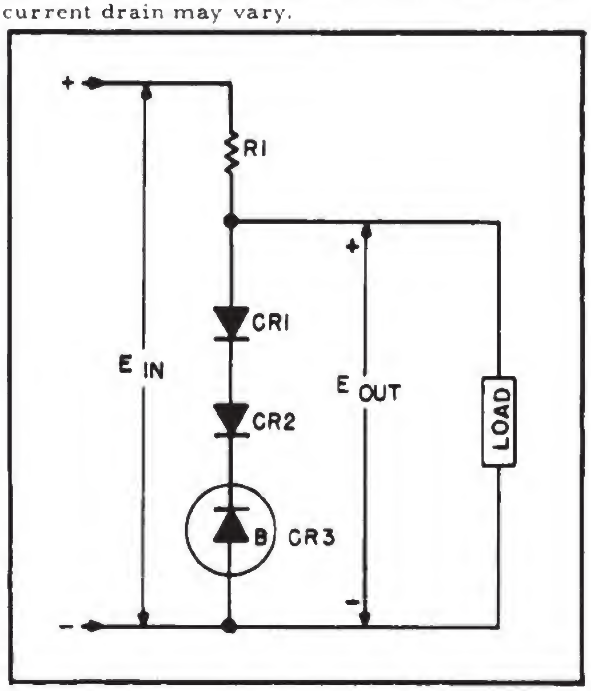

Figure 30-45 - Breakdown diode temperature compensated voltage regulator.

Two diodes are used in this circuit because the temperature-coefficient of resistance of each is half that of the breakdown diode. Forward-biased diodes are used because of the very low voltage drop across them. Thermistors, or other temperature-sensitive elements, can also be used.

Q26. For what reason is the additional diode employed in double diode stabilization?

Q27. Why is double diode stabilization more desirable than forward biased single diode stabilization?

FREQUENCY LIMITATIONS

30-31. Alpha Cut-Off Frequency

The current gain, voltage gain, and power gain of a transistor amplifier are directly proportional to the short circuit forward current amplification factor. This factor is important in transistor electronics and is referred to as alpha$_{fb}$ (α_{fb}). With the subscript "f" being the current from the emitter to the collector in the forward direction, and the subscript "b" indicating that a common-base configuration is being used. The short circuit forward current amplification factor is the ratio of the output current to the input current, measured with the output short circuited (ac). If alpha$_{fb}$ of a given transistor is high, the B will be correspondingly high; therefore the following discussion is concerned only with alpha$_{fb}$.

To obtain maximum gain in a given amplifier, it is necessary to use a transistor having a high-valued short circuit forward current amplification factor (alpha$_{fb}$). Most transistors have an alpha$_{fb}$ ranging from 0.940 to 0.985 with the average value of 0.96. Regardless of the average value of alpha$_{fb}$, this factor will decrease as emitter current increases. In a typical transistor the ratio of collector current to emitter current will remain relatively constant until a high value of emitter current is reached; then the collector current will increase less rapidly as the emitter current increases. This indicates a reduction in the current amplification factor at high values of emitter current. This phenomenon is most pronounced in power amplifiers that draw heavy emitter current.

A more important limitation placed on alpha$_{fb}$ is that of frequency. Like all amplifying devices the transistor has a maximum operating frequency. The frequency limitation of transistors will vary from one type to another with some types designed to operate as high as 1000 Mc.

The short circuit current gain alpha$_{fb}$ is measured (in most cases) at a frequency of 1 kc. The ALPHA CUT-OFF FREQUENCY is defined as that frequency at which alpha$_{fb}$ drops to 0.707 of its 1 kc value. The alpha cut-off frequency is designated as $f_{\alpha co}$. For the common-emitter configuration the beta cut-off frequency is designated f_{Bco}.

30-32. Interelement Capacitances

Another factor which makes the gain of a transistor frequency dependent are the INTERELEMENT CAPACITANCES within the transistor.

Figure 30-46 shows the interelement capacitances associated with the transistor. The

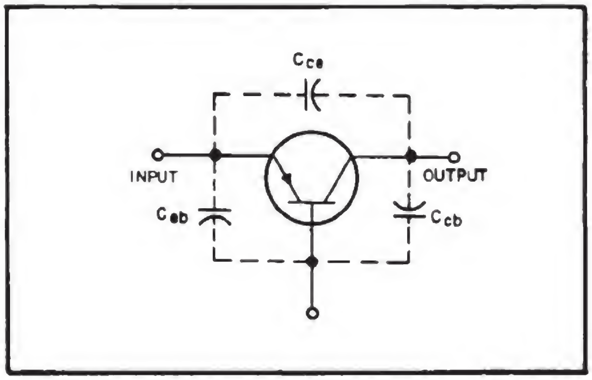

Figure 30-46 - Effective interelement capacitances of the transistor.

capacitances are shown externally; however, the actual capacitance effects are produced by the PN junctions within the transistors. Because the width of the PN junctions vary in accordance with the voltages across them and the current flow through them, the capacitances values also vary. For example, the variation of collector-base capacitance (C_{cb}) with collector voltage and emitter current is shown in A and B, Figure 30-47.

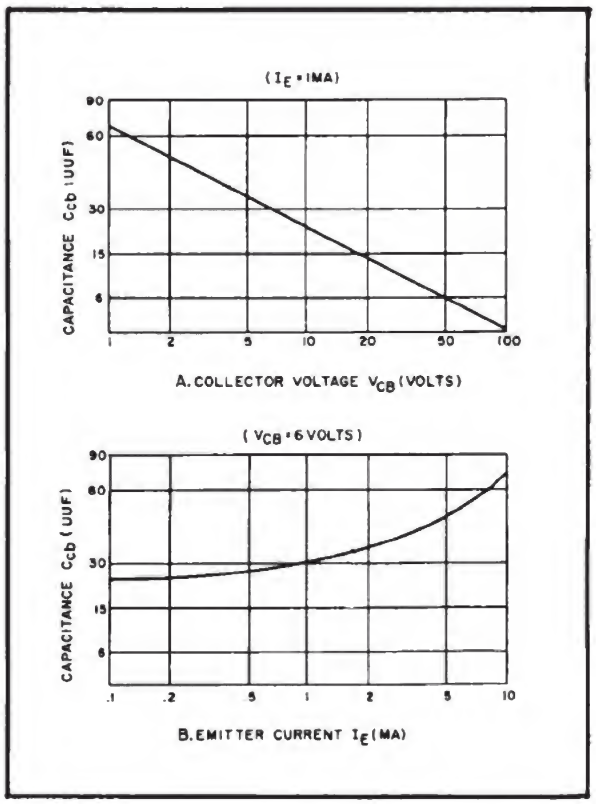

Figure 30-47 - Variation of C_{cb} with V_C and I_E.

A26. To compensate for collector current variations when the circuit is operated above 50° C.

A27. Single diode stabilization does not compensate for I_{CBO}.

The increase in the width of the PN junction between base and collector, as the reverse bias voltage (V_{CB}) is increased, is reflected in lower capacitance values. This phenomenon is equivalent to spreading apart the plates of a capacitor so that lower capacitance results.

An increase in emitter current, most of which flows to the collector through the base-collector junction, increases the collector-base capacitance (C_{cb}). The increased current through the PN junction may be considered as effectively reducing the width of the PN junction. This phenomenon is equivalent to reducing the distance between the plates of a capacitor so that increased capacitance results.

In high-frequency amplifier applications, the collector-base capacitance causes positive feedback that may result in oscillation. External circuits must be used to prevent oscillation.

The average value of collector-base capacitance (C_{cb}) may vary from 2 pf for high frequency transistors, to 50 pf for low frequency (audio) transistors.

The collector-emitter capacitance (C_{ce}) of a transistor, also caused by the PN junction, normally is greater than that of the collector-base capacitance and also varies with emitter current and collector voltage.

The effect of the variation of interelement capacitance with current and voltage is utilized to frequency modulate oscillators.

Because the collector-emitter capacitance is greater than the collector-base capacitance, common-base amplifiers have a better high frequency response than common-emitter amplifiers.

The emitter-base capacitance (C_{eb}), although very high because of normal forward bias and resultant small width of the PN junction, does not offer many problems in amplifier design because it is normally shunted by a low input resistance.

Figure 30-48 shows an RC coupled amplifier with interelement capacitances inserted. The high frequency response is limited by C_o and C_i. The stray capacitance and the capacitance of the output impedance are represented by C_o. The capacitive effect of the input impedance is represented by C_i. Since the reactances of C_o and C_i decrease as the frequency increases, the gain falls off as the frequency increases.

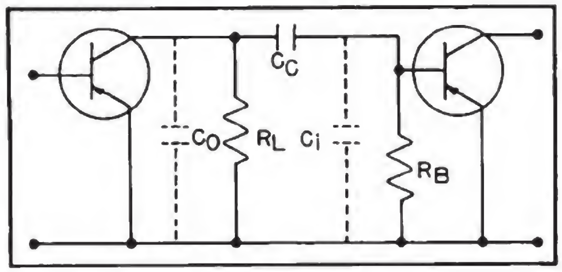

Figure 30-48 - Interelement capacitances in relation to external circuit.

The low frequency response is limited by the time constant of capacitor C_c and resistor R_B. This time constant must be long in comparison to the lowest frequency to be amplified.

Q28. What causes C_{cb} to decrease as V_C is increased?

Q29. In a CB configuration, C_{cb} is shunted across the load. Which of the interelement capacitances is shunted across the load in the CE configuration?

NEUTRALIZATION AND UNILATERALIZATION

A UNILATERAL electrical device is one which transmits energy in one direction only. The transistor is not a unilateral device. Every transistor, due to interelement capacitances, has an internal feedback mechanism which feeds back voltage from the output to the input. In each transistor configuration, the feedback voltage aids the input voltage and is therefore a positive feedback. If the positive feedback voltage is large enough, the amplifier will oscillate. At low frequencies (audio frequencies), the voltage feedback is low; therefore, special precautions to prevent oscillation are not usually required. High frequency amplifiers with tuned coupling networks are susceptible to oscillations so that precautions must be taken.

The effect of positive or negative voltage feedback on the input circuit of an electrical device is to alter its input impedance. Normally, both the resistive and the reactive components of the input impedance are affected. The change in the input impedance of a transistor caused by the internal feedback, can be eliminated by using an external feedback circuit. If the external feedback circuit cancels both the resistive and the reactive changes in the input circuit, the transistor amplifier is considered to be UNILATERALIZED. If the

external feedback circuit cancels only the reactive changes in the input circuit, the transistor amplifier is considered to be NEUTRALIZED. Actually, neutralization is a special case of unilateralization. In either case, the external feedback circuit prevents oscillation in the amplifier.

30-33. CB Amplifier, Unilateralized

Figure 30-49 shows a tuned common-base amplifier. The dc biasing circuits are not shown. Transformer T_1 couples the input signal to the amplifier. Capacitor C_1 and the transformer T_1 secondary winding form a parallel resonant circuit. Transformer T_2 couples the output of the amplifier to the following stage. Capacitor C_2 and the transformer T_2 primary winding form a parallel resonant circuit.

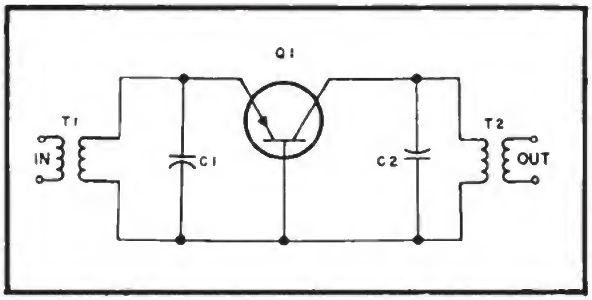

Figure 30-49 - CB amplifier, tuned.

The internal elements of the transistor that may cause sufficient feedback and oscillation are shown in dashed lines in Figure 30-50.

Resistor r'_b represents the resistance of the bulk material of the base and is referred to as the BASE-SPREADING RESISTANCE. Capacitor C_{cb} represents the capacitance of the base-collector junction. Resistor r_c represents the resistance of the base-collector junction. This resistance is very high in value because of the reverse bias on the base-collector junction. At very high frequencies, capacitor C_{cb} effectively shunts resistor r_c.

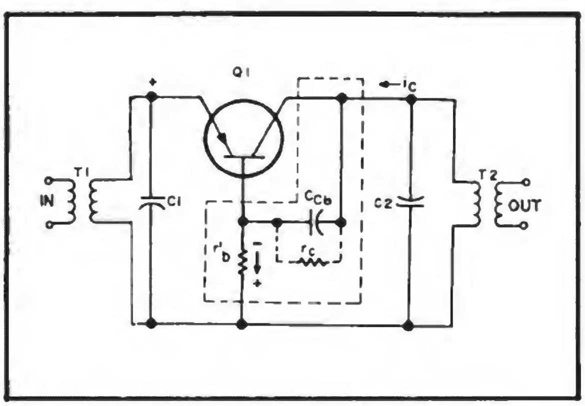

Figure 30-50 - Internal feedback elements.

Assume that the incoming signal aids the forward bias (causes the emitter to go more positive with respect to the base). Collector current increases in the direction shown. A portion of the collector current passes through capacitor C_{cb} and through resistor r'_b in the direction shown, producing a voltage with the indicated polarity. The voltage across resistor r'_b aids the incoming signal and therefore represents a positive feedback that may cause oscillation.

To overcome the possibility of oscillation, the external circuit (consisting of resistors R_{N1} and R_{N2} and capacitor C_N) can be added to the circuit, as shown in Figure 30-51. Resistors R_{N1} and R_{N2} and capacitor C_N correspond to resistors r'_b and r_c and capacitor C_{cb} respectively. The need for resistor R_{N2} depends on the frequency being amplified; the higher the frequency, the less the need for resistor R_{N2}.

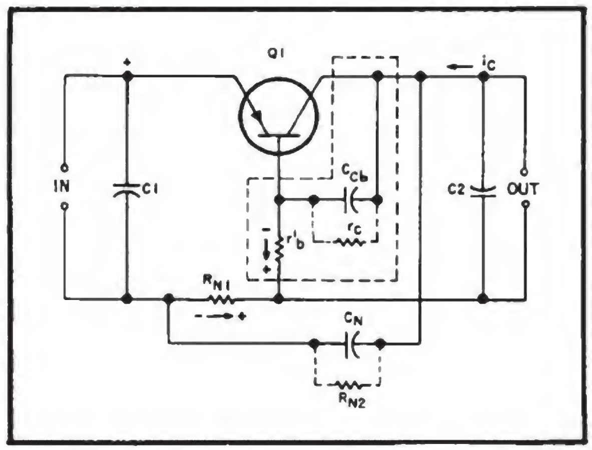

Figure 30-51 - Unilateralized CB amplifier.

When the incoming signal aids the forward bias, the collector current increases. A portion of the collector current passes through capacitor C_{cb} through resistor r'_b in the direction shown, producing a voltage with the indicated polarity. A portion of the collector current also passes through capacitor C_N and through resistor R_{N1}, producing a voltage with the indicated polarity. The voltages produced across resistors r'_b and R_{N1} are opposing voltages. If the voltages are equal, no positive or negative feedback from output circuit to input circuit occurs. The amplifier is considered to be unilateralized.

Q30. What type of feedback is developed across r'_b?

Q31. How is unilateralization different from neutralization?

A28. Increasing V_C causes the electrostatic field of the CB junction to increase. Increasing the width of the field is the same as moving the plates of a capacitor further apart, thus, capacitance decreases.

A29. C_{ce} shunts the load in the CE configuration.

A30. Positive or regenerative feedback.

A31. Unilateralization cancels BOTH resistive and reactive changes in the input circuit.

30-34. Common-Emitter Amplifier, Partial Emitter Degeneration

A common-emitter amplifier using partial emitter degeneration to unilateralize the amplifier is shown in Figure 30-52. Capacitor C_N and resistors R_{N1} and R_{N2} form a unilateralizing network. Transformer T_1 couples the input signal to the base-emitter circuit. Resistor R_1 forward biases the base-emitter circuit. Capacitor C_1 prevents shorting of base bias voltage by the transformer T_1 secondary. Transformer T_2 couples the output signal to the following stage. Capacitor C_2 and the transformer T_2's primary winding form a parallel resonant circuit. Capacitor C_3 blocks battery dc voltage from resistor R_{N1} and couples a portion of the collector current (i_{c2}) to the emitter. Inductor L_1 (an rf choke) prevents ac shorting to ground of resistor R_{N1} through capacitor C_3.

When the input signal aids the forward bias, collector current i_c increases in the direction shown. Internally, a portion of the collector current is coupled to the base-spreading resistance through the collector-base junction capacitance. The voltage developed across the base-spreading resistance aids the incoming signal and constitutes a positive feedback. To offset the positive feedback, a portion of the collector current (i_{c1}) is directed through resistor R_{N1} and the parallel combination of capacitor C_N and resistor R_{N2}. The voltage developed across resistor R_{N1} is a degenerating voltage equal and opposite to that developed across base-spreading resistance. The net voltage feedback to the input circuit is zero, and the amplifier is thus unilateralized. The values of capacitor C_N and resistors R_{N1} and R_{N2} depend on the internal values of the collector-base junction capacitance, the base-spreading resistance, and the collector resistance, respectively.

Q32. What would be the polarity of the charge on neutralizing capacitor C_N in Figure 30-51 with the circuit operating?

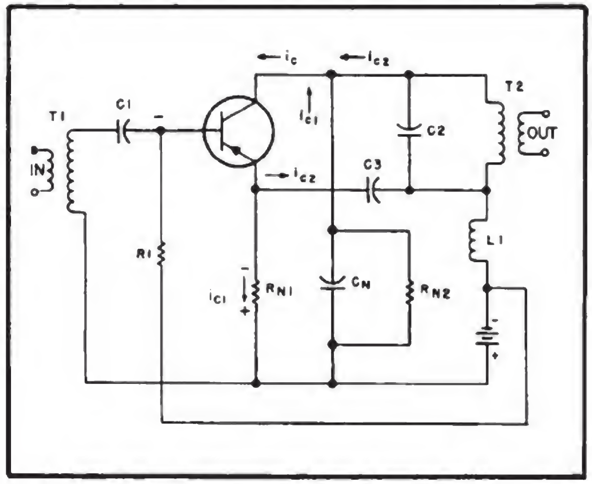

Figure 30-52 - CE amplifier, partial emitter degeneration.

FREQUENCY COMPENSATION

30-35. High Frequency Shunt Compensation

For shunt compensation (Figure 30-53), inductor L_1 is added in series with load resistor R_L. This compensates for the shunting effect of output impedance capacitance C_o and input impedance C_i. The capacitive reactance of output impedance capacitance C_o and input impedance capacitance C_i shunt inductor L_1 and load resistor R_L. Capacitor C_c is practically a short circuit at high frequencies. Since inductor L_1 (in series with resistor R_L) and capacitors C_i and C_o form a parallel resonant circuit having a very broad response, this type of compensation may also be called shunt peaking. The resonant peak of this parallel combination maintains a practically uniform gain in the high frequency range. In the uncompensated circuit the high frequency gain was reduced by the capacitive reactances of capacitors C_o plus C_i. The impedance of the parallel resonant circuit has a value that is approximately the same as load resistor R_L. When the frequency increases, the decrease in the capacitive re-

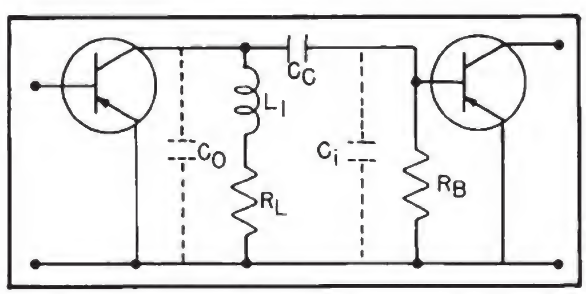

Figure 30-53 - Shunt compensation.

actances of C_o plus C_i is exactly compensated by the increase in the inductive reactance of inductor L_1. The frequency response is increased, extending the flatness of the response curve over a much higher range of frequencies.

30-36. High Frequency Series Compensation

For SERIES COMPENSATION, Figure 30-54, inductor L_2 is added in series with capacitor C_c. Considering capacitor C_c as a short circuit at high frequencies, inductor L_2 and capacitor C_i form a series resonant circuit at very high frequencies. As very high frequencies are approached, the series circuit approaches resonance. Thus, current flow through capacitor C_i increases as the frequency increases. The capacitive reactance of capacitor C_o decreases the voltage across load resistor R_L as the frequency increases. Since the current flow through capacitor C_i increases with frequency it compensates for the decrease in voltage across load resistor R_L. The frequency response is approximately the same as for the shunt peaking coupling circuit. The high frequency gain of a wide-band amplifier using series peaking is about 50 percent greater than that using shunt peaking.

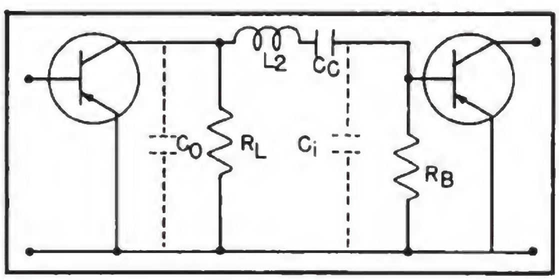

Figure 30-54 - Series compensation.

30-37. High Frequency Series-Shunt Compensation

For SERIES-SHUNT COMPENSATION, Figure 30-55, the series-compensation and the shunt-compensation coupling circuits are combined. This type of coupling is called combination compensation or combination peaking. The qualitative analysis discussed in shunt compensation and series compensation apply to this circuit. The frequency response is approximately the same as that of the shunt peaking or the series peaking coupling circuits. The high frequency gain of the combination peaking coupling circuit is approximately 80 percent greater than the series peaking coupling circuit.

30-38. Low Frequency Compensation

On the low-frequency end of the frequency response range, the input and the output capaci-

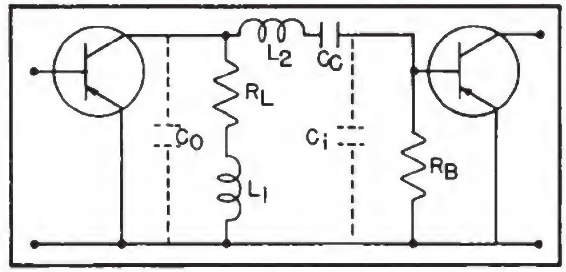

Figure 30-55 - Series-shunt compensation.

tances of the transistors have no effect on the frequency response. The low-frequency response is limited by capacitor C_c (Figure 30-56) and resistor R_B. The time constant ($R_B C_c$) must be large to prevent the low frequency response from falling off and to prevent phase distortion. The loss of gain at low frequencies is minimized by adding a compensating filter in series with load resistor R_L. The compensating filter consists of resistor R_F and capacitor C_F. The filter increases the collector load impedance at low frequencies and compensates for the phase shift produced by capacitor C_c and resistor R_B. For high frequencies, capacitor C_F is practically a short circuit. Thus, at high frequencies, the collector load impedance consists only of resistor R_L. As the frequency decreases, the reactance of capacitor C_F increases. For very low frequencies capacitor C_F is practically an open circuit. Thus, at very low frequencies the collector load impedance consists of resistor R_L and resistor R_F. This combination extends the frequency response curve over a much lower range of frequencies. The gain and phase response becomes more uniform over the low frequency range of the frequency response curve.

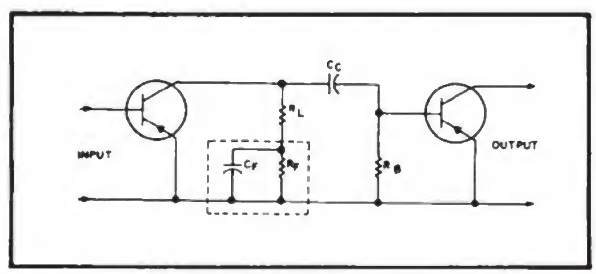

Figure 30-56 - Low frequency compensation.

30-39. Simultaneous High- and Low-Frequency Compensation

Figure 30-57 shows a typical wide-band amplifier (30 cps to 4 Mc) using both high- and low-frequency compensation. The high- and the low-frequency compensating circuits operate independently and do not interfere with each

A32. Negative on the plate connected to the collector of the transistor and positive on the plate connected to the plus terminal of the battery.

---

other. At low frequencies the series reactance of inductor L_1 is very small and has no effect on the collector load impedance. The series reactance of inductor L_2 is also very small and has no effect on the input circuit. The reactances of output impedance capacitance C_o (not shown) and input impedance capacitance C_i (not shown) are so large that they have no effect at low frequencies. This means that the combination peaking coupling has no effect at low frequencies. Similarly, at high frequencies, the reactance of capacitor C_F is very small and is practically a short circuit. Resistor R_F has no effect on the collector load impedance at high frequencies. Thus, compensating filter $R_F C_F$ has no effect at high frequencies.

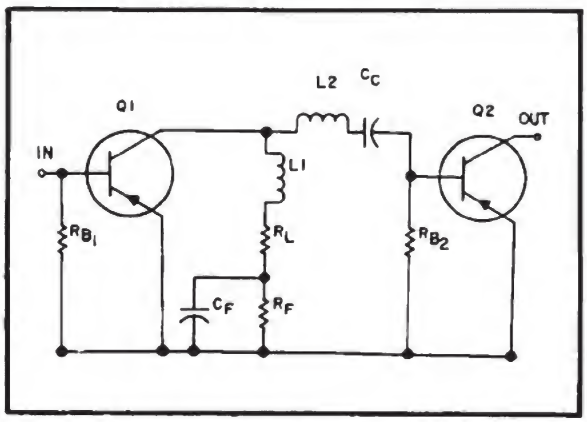

Figure 30-57 - High and low frequency compensation.

Q33. Does shunt compensation or series compensation yield the greatest gain at high frequencies?

Q34. In a circuit employing low frequency compensation (such as Figure 30-57), what would be the effect on HIGH frequency operation if capacitor C_F were to become shorted?

SERVICING TECHNIQUES

Transistors may be damaged beyond repair by applying the incorrect polarity to the collector circuit or excessive voltage to the input circuit or by careless soldering techniques that overheat the transistor. Also the low-voltage electrolytic capacitors used in transistor circuits are easily damaged by reversed polarity connections or excessive voltage.

30-40. Soldering Techniques

Because of the small physical size of transistors and their associated circuit components, small-sized tools are used. Small cutting pliers and needle-nose pliers are more useful than the conventional sizes. Narrow-blade screw drivers are more useful than larger, conventional types. A sharp-pointed, thin metal probe is also useful for cleaning solder from small openings or areas. Soldering is performed more satisfactorily with a small low-wattage soldering iron (35 to 40 watts) having a narrow point or wedge as its tip.

Always ground the frame of the soldering iron or gun to the chassis when soldering around transistors or transistor circuitry. Enough leakage voltage is present in most irons to cause damage to the transistor. Soldering with an ungrounded iron to a transistor and its associated circuitry held in one's hand may damage the transistor because the body has considerable capacitance to ground; the charging of this capacitance through the transistor can cause permanent damage. The junction is made very thin in order to operate at very high frequencies and is easily damaged by excess current. The same precaution (grounding the frame or chassis) is followed in applying any ac operated test equipment.

When soldering transistor leads, the terminal lead is held with needle-nose pliers (or a heat shunt) positioned between the transistor body and the lead end. This arrangement allows the heat to travel into the pliers (or shunt), thereby diverting it away from the transistor. To make sure that all the heat is drawn away from the transistor, the pliers should remain securely on the lead for a short interval after removing the iron. Transistor leads should be relatively long and the soldering operation should be as short as possible. Low-temperature rosin core solder is the proper type to use.

Where transistor leads are stiff they may be plugged into appropriate type sockets. In this case the socket terminals should be soldered only when the transistor is out of the socket.

Transistor components are small and the connecting wires of resistors, capacitors, and coils are easily broken. These wires should be handled carefully both in installations and in replacements.

30-41. Testing Transistor Circuits

Transistors are biased in the forward, or low resistance direction, of the base-emitter

input circuits. These circuits are particularly vulnerable to any excess voltage over their rated value. This excess voltage can cause a current to flow that will overheat the transistor junction and permanently damage it. Proper operation depends on the crystal lattice structure and the impurity atoms that are present. Heat will distort the lattice and affect the behavior of the impurity atoms so that normal transistor action is seriously impaired. Excessive collector current will generate more heat in the relatively high impedance of the collector circuit and damage the transistor. Maximum allowable collector current depends on the ambient temperature. If the ambient temperature goes up, the maximum allowable collector current must be decreased (derated).

Both the value and the polarity of voltages to be applied to transistor circuits should be carefully checked before they are applied. It is important to know the type of transistor being used. The NPN types require positive collector voltages and negative emitter voltages; whereas, the PNP types require negative collector voltages and positive emitter voltages. Transistors should be inserted properly in the sockets before voltages are applied in order to avoid the possibility of transient currents (momentary pulses) which might damage a transistor. Similarly, power to the unit should be disconnected before removing transistors from sockets. If the magnitude of the collector current is in doubt a milliammeter of suitable range should be connected in series with the collector terminal, and battery potential should be applied gradually by means of a potentiometer.

When signal generators are used in testing transistor circuits the magnitude of the signal should be limited to a low value, especially in low-level input stages. Indirect coupling, for example, (either capacitive or inductive) is preferred to direct coupling. The ungrounded (hot) lead from the output terminal of the signal generator can be terminated on the insulated portion of a capacitor or resistor in the circuit under test. Another method is to connect a coil to the output terminals of the signal generator and bring the coil into proximity with inductive elements of the circuit under test.

If an ohmmeter is used to check components in a transistor circuit, the range in use should not employ a battery of more than about 3 V. Higher voltages can damage the transistor. Electrolytic capacitors may give incorrect readings if the rule for polarities of the ohmmeter leads is not observed when connecting the leads to the capacitor terminals. The positive lead of the ohmmeter should connect to the positive terminal of the electrolytic capacitor, and the negative lead should connect to the negative terminal for correct indications.

The markings of the test leads do not necesarily indicate the polarity of the voltage on the leads when using a multimeter as an ohmmeter. The lead marked "common" may be either positive or negative with respect to the other leads depending on the design of that particular multimeter. Make sure by reference to the technical manual or by checking the polarity of the ohmmeter leads with a voltmeter. This is done by putting the common leads of the two meters together and the positive probe of the voltmeter on the ungrounded lead of the ohmmeter. The direction of deflection of the voltmeter needle will indicate whether the common lead of the ohmmeter is positive or negative with respect to the other leads. For example, when using the AN/PSM-4A as an ohmmeter, the lead marked "common" is positive with respect to the lead marked "volt-ohm-amps."

A transistor may be tested for its amplifying property by inserting it in a circuit like that of Figure 30-58 and applying a suitable input signal, preferably from an audio signal generator. The input and output voltages are then measured with a vacuum-tube voltmeter and the gain is calculated. This value is then compared with that of a transistor known to be in good condition.

There are many types of tests to be made on transistors, such as dc tests for opens, shorts, leakage and gain, plus the dynamic tests. The test circuit (Figure 30-58) is for PNP transistors, for use with NPN transistors the battery polarities are reversed by a switch (not shown).

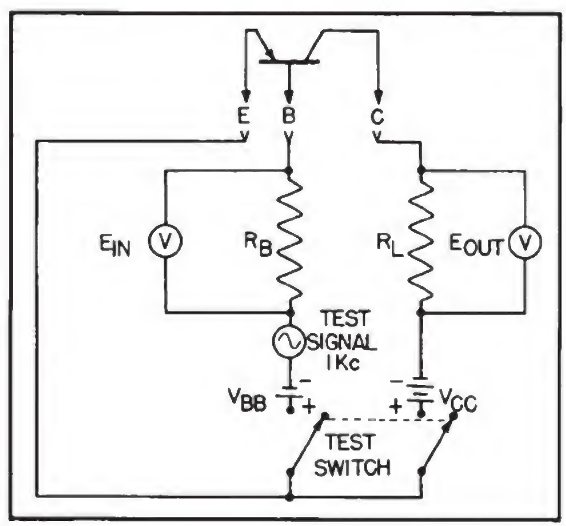

Figure 30-58 - PNP transistor tester.

The test switch allows insertion of the transistor in the test socket before application of

A33. Series compensation.

A34. Practically no effect. At high frequencies C_F appears as a short for normal operation.

---

the bias voltages. Since the test circuit is a common emitter configuration it is used specifically for dynamic beta measurement. The other tests are made with slightly revised test circuits.

When the output of a transistor radio receiver is distorted or weak, the first thing to check is the battery. Its voltage should be checked with a vacuum-tube voltmeter or high resistance volt-ohmmeter while the set is turned on. If the battery is weak, its voltage will be down 20 percent or more. Before replacing the battery the resistance between the battery clips should be checked. For example, if the resistance measured with an ohmmeter (requiring a battery of not over 3 V) is 8000 ohms and the manufacturer's limits are 6000 to 15,000 ohms, it is safe to insert a new battery. An alternative method of some manufacturers is to indicate the allowable current drain of the battery instead of the resistance between the battery clip with the battery removed. To obtain an indication of the current that the battery supplies to the set, a milliammeter can be connected in series with the battery and either clip.

ABBREVIATIONS AND SYMBOLS USED IN TRANSISTOR CIRCUITS

Semiconductor, General

BV........ Breakdown voltage
T_A........ Ambient temperature
T_{op}...... Operating temperature

Transistor

B, b...... Base electrode
C, c...... Collector electrode
C_{cb}...... Capacitance base-collector junction
C_{ib}...... Input capacitance (common-base)
C_{ie}...... Input capacitance (common-emitter)
C_{ob}...... Output capacitance (common-base)
C_{oe}...... Output capacitance (common-emitter)
E, e...... Emitter electrode
I_B........ Base current (dc)
i_b........ Base current (instantaneous)
I_C........ Collector current (dc)
i_c........ Collector current (instantaneous)
I_{CBO}..... Collector cutoff current (dc), emitter open
I_{CEO}..... Collector cutoff current (dc), base open
I_E........ Emitter current (dc)
i_e........ Emitter current (instantaneous)
R_B....... External base resistance
r_b........ Base spreading resistance, (sometimes labled r'_b)
r_i........ Input junction resistance
V_{BB}...... Base supply voltage
V_{BE}...... Base-to-emitter voltage (dc)
V_C....... Collector voltage (with respect to ground or common point.)
V_{CB}...... Collector-to-base voltage (dc)
V_{CE}...... Collector-to-emitter voltage (dc)
V_{ce}...... Collector-to-emitter voltage (rms)
v_{ce}...... Collector-to-emitter voltage (instantaneous)
V_{CE}(sat).. Collector-to-emitter saturation voltage
V_{EBO}.... Emitter-to-base voltage (static)
V_{CC}...... Collector supply voltage
V_{EE}...... Emitter supply voltage

EXERCISE 30

1. Give two ways in which a transistor and a vacuum tube are very different.
2. Define and explain the similarities between the elements of the electron tube and the transistor.
3. If the arrow head on the schematic symbol of a transistor points away from the base, what type of transistor is represented?
4. Explain why the emitter and the base can never be made of the same type of impurity (N or P) material in a single transistor.
5. Is the net charge on the N side of the junction negative or positive? Explain.
6. Is the electrostatic field of the junction increased or decreased when forward bias is applied? Explain why the field is caused to change when bias is applied.
7. Why can't a semiconductor diode be used as an amplifier?
8. Describe the effect on the emitter-base and collector-base junction fields when normal operating bias is applied to a two junction transistor.
9. In a properly biased PNP type transistor which element has the most positive potential connected to it, base, collector, or emitter?
10. Name some of the factors involved in the transportation of a charge through a transistor.
11. Explain the difference between drift current and diffusion current.
12. What is a density gradient?
13. What happens to the carriers that recombine in the base region?
14. What would be the direction of base lead current if temperature increased to the point where I_{CBO} exactly equaled the base current I_B?
15. Determine the beta of a transistor if a 10 microampere change in base current causes a 2 ma change in collector current.
16. In a common emitter amplifier, does collector current increase or decrease when the input signal OPPOSES the emitter-base bias?
17. In a common emitter amplifier does the base resistance R_B develop the forward bias? Explain.
18. Explain why the CE configuration produces a 180° phase shift, but the CB and CC configurations do not.
19. Which quantity is held constant when developing a family of I_C-V_C characteristic curves for a CB configuration?
20. Determine the base current in a CE amplifier if R_B=19k ohms, V_{BB}=1.5 V, and the input junction resistance (BE) equals 1 k ohm.
21. Define the term cut-off current.
22. Explain the saturation region of a transistor.
23. Explain the breakdown voltage region.
24. Explain how the constant power dissipation curve is obtained.
25. What is the permissible operating region of a transistor?
26. Define and explain class A operation of a transistor amplifier.
27. Define and explain class B operation of a transistor amplifier.
28. Define and explain class C operation of a transistor amplifier.
29. Why is bias stabilization necessary?
30. What is meant by the term degradation factor?
31. What is the advantage of thermistor stabilization over resistor stabilization?
32. Give one main reason why diode stabilization is better than thermistor stabilization.
33. What are the main interelement capacitances in a transistor and how do they affect frequency response?
34. Define the alpha and beta cut-off frequency.
35. What is the purpose of unilateralization?
36. Explain a method of high frequency compenation.
37. Explain some of the main precautions to be taken when servicing transistorized printed circuit equipment.

774-639 O - 65 - 5

CHAPTER 31

RECEIVER PRINCIPLES

Many of the principles, circuits, and components discussed in previous chapters are directly applicable to radio receivers. A basic knowledge of them is therefore assumed in treating the material included in this chapter.

At the radio transmitter the carrier frequency is modulated by the desired signal, which may consist of coded characters, voice, music, or other types of signals. AMPLITUDE MODULATION (AM) occurs if the modulating signals cause the amplitude of the output to vary. Although there are other types of modulation, only AM receivers will be treated in this chapter.

The RF carrier wave with the modulating signal impressed upon it is transmitted through space in the form of an electromagnetic wave. As the electromagnetic wave passes across the receiving antenna, small ac voltages are induced in the antenna. These voltages are coupled into the receiver via the antenna coupling coil. The function of the receiver is to select the desired carrier frequency from those present in the antenna circuit and to amplify the small ac signal voltage. The receiver then separates the modulating signal (intelligence) and the carrier by the process of detection (rectification and the removal of the RF component) and amplifies the resultant audio signal to the proper magnitude to operate the audio reproducer (speaker or earphones).

The radio receivers discussed in this chapter are the TUNED-RADIO-FREQUENCY (TRF) receiver, and the SUPERHETERODYNE receiver.

The purpose of this chapter is to give an overall view of the manner in which a receiver converts the electromagnetic wave, received by the antenna, into the type of energy that the human ear interprets—sound. This is accomplished by separating the receiver into blocks with each block representing a main section of the receiver. In this chapter the purpose and action of each block or section will be discussed briefly. Each section will then be the subject of a complete chapter which will discuss that section, and its relation to the entire receiver in detail.

31-1. Origin of Receiver Input

It has been stated that the purpose of a receiver is to convert the electromagnetic wave, from the transmitter, into energy usable by the human ear. Before entering into a discussion of the manner in which the receiver accomplishes this, the reader may find a brief review of the origination of the transmitted wave advantageous. Figure 31-1 illustrates the block diagram of a basic transmitter and the nature of the input and output energy of each section.

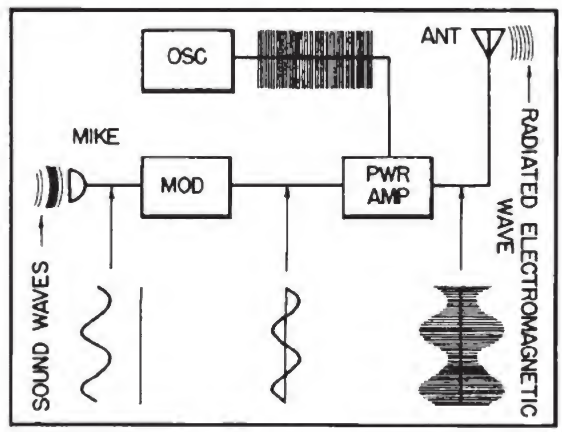

Figure 31-1 - Basic transmitter.

The input to the transmitter is in the form of sound waves which are produced by voice, musical instruments, etc. These sound waves range in frequency from about 20 cycles per second to around 20,000 cycles per second. The microphone converts the sound waves into an electrical signal which varies in frequency and amplitude in accordance with the original sound. Since the electrical signals from the microphone are very weak they are fed to the modulating section (audio amplifier) which increases the amplitude to a level suitable to be used as an input to the power amplifier.

The power amplifier, in the RF unit has two inputs; one is the audio signal from the modulator and the other is the constant amplitude radio frequency signal from the oscillator. In the broadcast band, the channel frequencies for various transmitter's will range from 535 kc to 1605 kc. The output of the oscillator is called the CARRIER frequency. In the power amplifier, the audio signal (intelligence) is impressed on

the carrier. The output of the power amplifier is a modulated RF signal, which is then fed to the antenna.

The antenna radiates the modulated RF signal in the form of electromagnetic waves. This electromagnetic wave will now be used as the input to the receiver.

Q1. What is the function of the microphone?

Q2. What are the essential functions that a receiver must perform?

Q3. The sensitivity ratings of two receivers are being compared. Receiver X has an output of one watt with an input signal of 50 micro-volts. Receiver Y has an output of one watt with an input signal of 25 micro-volts. Which receiver has the highest sensitivity?

31-2. Functions of a Receiver

A receiver must perform certain basic functions in order to be useful. These functions, in order of their performance are: RECEPTION, SELECTION, DETECTION, A. F. AMPLIFICATION, and REPRODUCTION.

RECEPTION involves having the transmitted electromagnetic wave pass through the receiver antenna in such a manner as to induce a voltage in the antenna.

SELECTION involves being able to select a particular station's frequency from all the transmitted signals that happen to be induced in the receiver's antenna at a given time.

DETECTION is the action of separating the low frequency intelligence from the high frequency carrier.

AF AMPLIFICATION involves amplifying the low frequency intelligence (audio in the case of a radio) to the level required for operation of the reproducer.

REPRODUCTION is the action of converting the electrical signals to sound waves which can then be interpreted by the ear as speech, music, etc.

The ability of a receiver to reproduce the signal of a very weak station is a function of the receiver's SENSITIVITY. In other words, the weaker a signal that can be applied to a receiver, and still achieve the same value of signal output, the better is that receiver's sensitivity rating.

The ability of a receiver to select and reproduce a desired signal from among several closely spaced stations, or from among interferring frequencies, is determined by the receiver's SELECTIVITY. In other words, the better a receiver is at differentiating between desired and undesired signals, the better is the receiver's selectivity rating.

31-3. Block Diagram of Simple Receiver

Figure 31-2 shows the block diagram of a simple receiver which will perform all the functions required of a receiver.

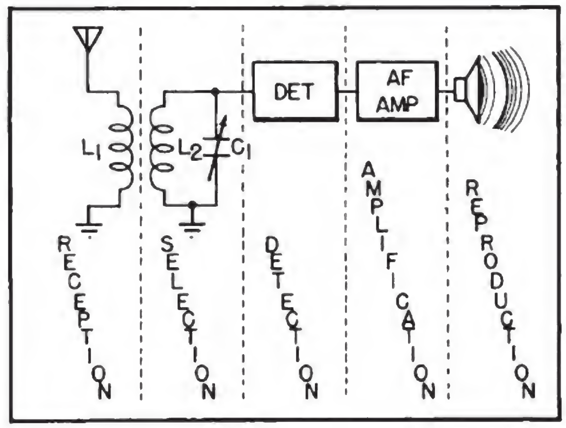

Figure 31-2 - Simple receiver block diagram.

Figure 31-2 also illustrates the functions performed by the various sections of the receiver. The input to the receiver is the electromagnetic wave propagated from the antenna of the transmitter. This wave will pass through the antenna of the receiver and induce a small ac voltage. The section of the receiver formed by the antenna and L_1 perform the function of reception. L_1 is the primary of the input transformer and the voltage induced in L_1 is coupled to the secondary, L_2. L_2 and C_1 form a tuned circuit with C_1 being variable to permit tuning across the broadcast band. Thus, the tuned input circuit performs the function of selecting a specific frequency from among those present in the antenna circuit. The output of the tuned circuit is a modulated RF signal.

This modulated RF signal is then fed to the detector circuit where the function of detection (rectification and filtering) is performed. The output of the detector circuit is a weak audio signal. The audio signal from the detector is too weak to satisfactorily operate a speaker; therefore, it is fed to an audio frequency amplifier to increase its amplitude. The output of the AF amplifier is fed to the speaker, which performs the function of reproduction, or converting the electrical signals back to the form of the original input to the transmitter (in this case, sound waves).

TRF RECEIVER

The tuned radio frequency receiver, generally known as the TRF receiver, consists of

one or more RF stages, a detector stage, one or more AF stages, a reproducer, and the necessary power supply. A block diagram of a TRF receiver is shown in Figure 31-3. The waveforms that appear in the respective sections of the receiver are shown above the block diagram.

The amplitude of the AM signal at the input of the receiver is relatively small because it has been attenuated in the space between the transmitter and the receiver. It is composed of the carrier frequency and two sideband frequencies. The RF amplifier stages amplify the waveform, but they do not change its basic shape if the circuits are operating properly. The detector rectifies and removes the RF component of the signal. The output of the detector is a weak signal made up only of the modulation component, of the incoming signal. The AF amplifier stages following the detector increase the amplitude of the AF signal to a value sufficient to operate the loudspeaker or earphones.

31-4. RF Section

The ANTENNA-GROUND SYSTEM serves to introduce the desired signal into the first RF amplifier stage via the antenna coupling transformer. For best reception the resistance of the antenna-ground system should be low. The antenna should also be of the proper length for the band of frequencies to be received, and the antenna impedance should match the input impedance of the receiver. The gain of most commercial receivers; however, is generally sufficient to make these values non-critical.

The RF AMPLIFIERS in the TRF receiver have tunable tanks in the input circuits. Thus, the receiver may be tuned so that only one RF signal within its tuning range is selected for amplification. When the tank is tuned to the desired frequency, it resonates and produces a relatively large circulating current. The input of the RF amplifier then receives a relatively signal voltage at the resonant frequency and minimum signal voltage at all other frequencies.

It must be remembered that the amplitude modulated signal as it leaves the transmitter, with a single audio frequency as the intelligence, is actually composed of energy at three distinct frequencies. These frequencies are the carrier frequency, and two (upper and lower) sideband frequencies. The separation between the maximum limits of the upper and lower sideband frequencies constitutes the bandwidth of the transmitted signal. The separation between either sideband and the carrier is equal to the intelligence frequency. In the case of radio, the energy content of the sideband frequencies is a direct function of the percent of modulation (intensity) of the audio frequencies.

Although broadcast band receiver's cover a frequency range from 535 kc to 1605 kc, the bandwidth of any specific station is only a small percentage of this operating frequency. Station bandwidths range from 10 kc to 14 kc depending on whether the station operates in the upper or lower portion of the frequency range (high frequency stations have the lower bandwidths). Therefore, the RF amplifiers and their associated interstage coupling networks must be capable of passing a band of frequencies as wide, or wider, than the station bandwidth.

The sensitivity and selectivity of a TRF receiver is improved by increasing the number of RF stages. However, there is a practical limit to the number of stages that may be added (the reasons behind the advantages and limitations will be explained in Chapter 32). The tuned tanks of each RF stage, in a TRF receiver, must be tuned to the same frequency. Therefore, when changing frequency all of the tanks must be changed to the new frequency. When two or more RF stages are used, the tuning capacitors (coil cores if inductively tuned) are ganged on the

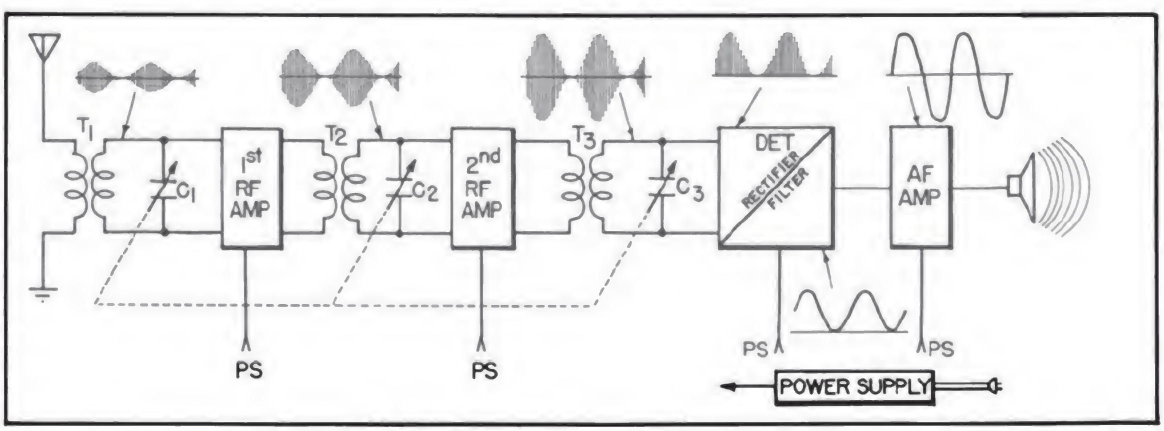

Figure 31-3 - Block diagram of a TRF receiver and waveforms.

A1. To convert sound waves into electrical signals.

A2. Reception, selection, detection, amplification, and reproduction.

A3. Receiver Y.

---

same shaft. This is indicated by a dotted line connecting the tanks (Figure 31-3).

Q4. What effect do the RF amplifiers have on the waveshape of the signal?

31-5. Detector Section

The process of removing the intelligence component of the modulated waveform from the RF carrier is called DETECTION or DEMODULATION. In the amplitude modulation system, the audio or intelligence component causes both the positive and the negative half cycles of the RF signal to vary in amplitude. The function of the detector is to rectify the modulated signal. A suitable filter eliminates the remaining RF pulses and passes the audio component on to the AF amplifiers. In Figure 31-3 the detector section is divided into two parts, the rectifier, and the filter. The waveform of the rectifier portion of the detector section indicates the positive half cycles of the signal remaining after rectification. The waveform for the filter portion of the detector section indicates the audio component remaining after the RF pulses have been filtered out.

The various types of detector circuits will be explained in detail in Chapter 36.

31-6. Audio Frequency Section

The function of the AF section of a receiver is to further amplify the audio signal. In most cases the amount of amplification that is necessary depends on the type of reproducer used. If the reproducer consists of earphones, only one stage of amplification may be necessary. If the reproducer is a large speaker or other mechanical device requiring a large amount of power, several stages may be necessary. In most receivers the last stage is operated as a power amplifier. The reproducer converts the audio signal to sound waves.

31-7. Power Supply Section

The power supply section of the receiver converts the ac energy from the power line into the proper value of dc energy required by the individual sections of the receiver.

31-8. Characteristics of the TRF Receiver

The principal disadvantage of the TRF receiver is that its selectivity, or its ability to separate signals, does not remain constant over its tuning range. As the set is tuned from the low-frequency end of its tuning range to the high frequency end, its selectivity decreases.

Also, the amplification, or gain, of a TRF receiver is not constant over the tuning range of the receiver. The gain depends on RF transformer gain, which increases with frequency. In order to improve the gain at the low-frequency end of the band, RF transformers employing high-impedance (untuned) primaries are designed so that the primary inductance will resonate with the primary distributed capacitance at some frequency slightly below the low end of the tunable band. Thus, the gain is good at the low end of the band because of the resonant buildup of primary current. The near-resonant condition of the primary at the low end more than offsets the effect of reduced transformer action. However, the shunting action of the primary distributed capacitance lowers the gain at the high-frequency end of the band. To make up for the resultant poor gain at the high end of the band, a small capacitor is connected between the plate and grid leads of adjacent RF stages to supplement the transformer coupling. At the low end of the band the capacitive coupling is negligible.

The superheterodyne receiver has been developed to overcome many of the disadvantages of the TRF receiver.

Q5. Does the detector section change the waveform of the audio component during the rectification process?

SUPERHETERODYNE RECEIVERS

The essential difference between the TRF receiver and the superheterodyne receiver is that in the former the RF amplifiers preceding the detector are tunable over a band of frequencies; whereas in the latter the corresponding amplifiers are tuned to one fixed frequency called the INTERMEDIATE FREQUENCY (IF). The principle of frequency conversion by heterodyne action is here employed to convert any desired station frequency within the receiver range to this intermediate frequency. Thus an incoming signal is converted to the fixed intermediate frequency before detecting the audio signal component, and the IF amplifier operates under uniformly optimum conditions throughout the receiver range. The IF circuits thus may be made uniformly selective, uniformly high in voltage gain, and uniformly of satisfactory bandwidth to contain all of the desired sideband components associated with the amplitude-modulated carrier.

The block diagram of a typical superheterodyne receiver is shown in Figure 31-4. Below corresponding sections of the receiver are shown

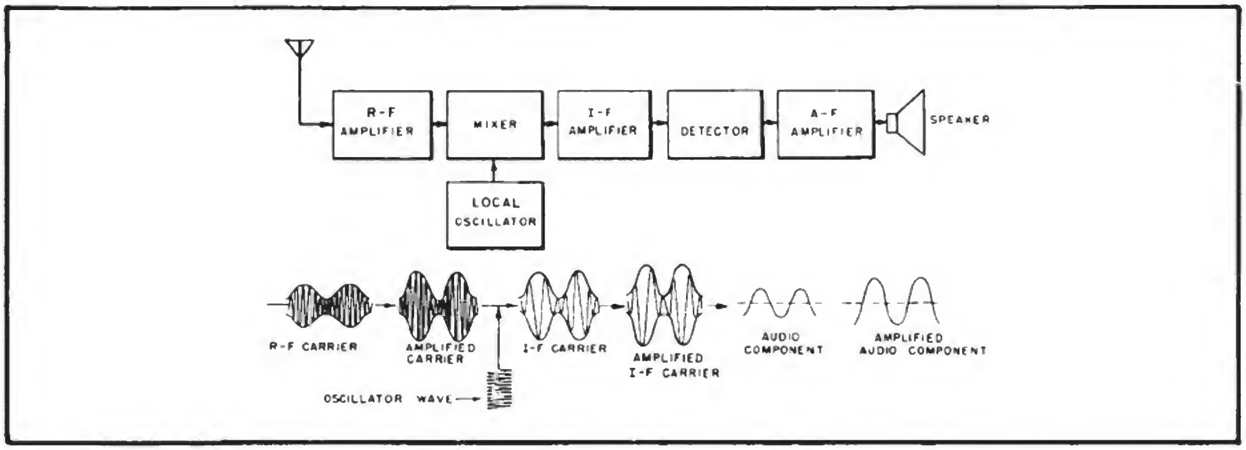

Figure 31-4 - Block diagram of a superheterodyne receiver and waveforms.

the waveforms of the signal at that point. The RF signal from the antenna passes first through an RF amplifier (preselector) where the amplitude of the signal is increased. A locally generated unmodulated RF signal of constant amplitude is then mixed with the carrier frequency in the mixer stage. The mixing or heterodyning of these two frequencies produces an intermediate-frequency signal which contains all of the modulation characteristics of the original signal. The intermediate frequency is equal to the difference between the station frequency and the oscillator frequency associated with the heterodyne mixer. The intermediate frequency is then amplified in one or more stages called INTERMEDIATE-FREQUENCY (IF) AMPLIFIERS and fed to a conventional detector for recovery of the audio signal.

The detected signal is amplified in the AF section and then fed to a headset or loudspeaker. The detector, the AF section, and the reproducer of a superheterodyne receiver are basically the same as those in a TRF set.

31-9. RF Amplifier

The RF stage amplifies the small ac voltages induced in the antenna by the electromagnetic wave from the station transmitter. Utilizing a tuned circuit between the antenna and the input of the RF amplifier permits selection of the desired station frequency from among the many present in the antenna.

If an RF amplifier is used ahead of the mixer stage of a superheterodyne receiver it is generally of conventional design. Besides amplifying the RF signal, the RF amplifier has other important functions. For example, it isolates the local oscillator from the antenna-ground system. If the antenna were connected directly to the mixer stage, a part of the local oscillator signal might be radiated into space. This signal could be picked up by a sensitive direction finder on any enemy ship. For this reason and others, Navy superheterodyne receivers are provided with at least one RF amplifier stage.

Also, if the mixer stage were connected directly to the antenna, unwanted signals, called IMAGES, might be received, because the mixer stage produces the intermediate frequency by heterodyning two signals whose frequency difference equals the intermediate frequency. (The heterodyne principle was treated in (Chapter 26).

The image frequency always differs from the desired station frequency by twice the intermediate frequency. Image frequency = station frequency $\pm$ (2 X intermediate frequency). The image frequency is higher than the station frequency if the local oscillator frequency tracks (operates) above the station frequency (Figure 31-5, A). The image frequency is lower than the station frequency if the local oscillator tracks below the station frequency (Figure 31-5, B). The latter arrangement is generally used for the higher frequency bands, and the former, for the lower frequency bands.

For example, if a superheterodyne receiver having an intermediate frequency of 455 kc is tuned to receive a station frequency of 1500 kc (Figure 31-5, A), and the local oscillator has a frequency of 1955 kc, the output of the IF amplifier may contain two interfering signals-one from the 1500-kc station and the other from an image station of 2410 kc (1500 + 2 X 455 = 2410 kc). The same receiver tuned near the low end of the band to a 590-kc station has a local oscillator frequency of 1045 kc. The output of the IF amplifier contains the station signal (1045-590 = 455 kc) and an image signal (1500-1045 = 455 kc). Thus the 1500-kc signal is an image heard simultaneously with the 590-kc station signal.

It may also be possible for ANY two signals having sufficient strength, and separated by the

A4. When operating properly the RF amplifiers will increase the amplitude of the waveshape, but change it in no other way.

A5. No. The waveform of the audio component will be almost identical to the waveform of the original modulating signal.

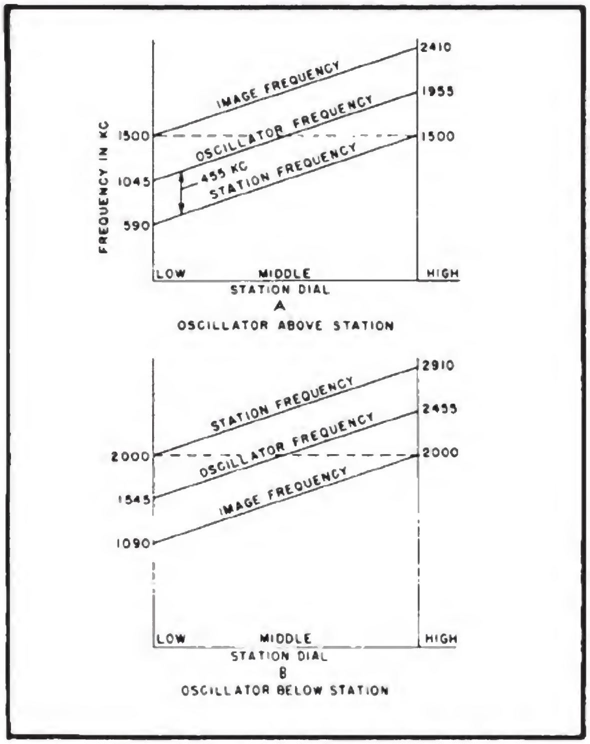

Figure 31-5 - Relation of image frequency to station frequency in a superheterodyne receiver.

intermediate frequency to produce unwanted signals in the reproducer. The selectivity of the preselector (RF stages preceding the mixer stage) tends to reduce the strength of these images and unwanted signals. However, there is a practical limit to the degree of selectivity obtainable in the preselector due to the fact that the RF stage must have a much wider bandwidth than the bandwidth of the desired signals (explained later).

The ratio of the amplitude of the desired station signal to that of the image is called the IMAGE REJECTION RATIO and is an important characteristic of a superheterodyne receiver.

The output of the RF amplifier stage is an amplified version of the modulated RF input signal.

Q6. What is the range of frequencies that the RF stage must be capable of tuning in a broadcast band superheterodyne receiver?

31-10. Local Oscillator

The function of the LOCAL OSCILLATOR stage is to produce a constant amplitude sine wave of a frequency which differs from the desired station frequency by an amount equal to the intermediate frequency of the receiver. The operation of the oscillator circuits is similiar to the operation of oscillator circuits previously discussed and will be explained in relation to the receiver in Chapter 33.

Although the oscillator may be operated either above or below the station frequency, in most broadcast band receivers the oscillator is operated above the station frequency. In order to allow selection of any frequency within the range of the receiver the tuned circuits of the RF stage and the local oscillator are variable. By using a common shaft, or ganged tuning, for the variable component of the tuned circuits, both circuits may be tuned in such a manner as to maintain the difference between the local oscillator frequency and the incoming station frequency equal to the receiver IF.

The output of the local oscillator is fed to the MIXER stage.

NOTE: The waveforms shown in Figure 31-4 are not shown to scale. In most receivers the amplitude of the oscillator waveform will be many times that of the modulated RF signal from the RF stage.

31-11. Mixer

The function of the MIXER stage is FREQUENCY CONVERSION by virtue of heterodyne action.

The input to the mixer consists of two signals: the modulated RF signal and the unmodulated local oscillator signal. The mixer then combines, or mixes, these two signals. As a result of this mixing action, the output of the mixer will contain four major frequencies plus many minor frequencies. The four major frequencies are: (1) The original signal frequency. (2) The local oscillator frequency. (3) The SUM of the signal and oscillator frequencies. (4) The DIFFERENCE of the signal and oscillator frequencies. The additional frequencies present are produced by combinations of the fundamentals and harmonics of the signal and oscillator frequencies. Of the frequencies present in the output of the mixer, only the difference frequency is used in amplitude modulated broadcast band receivers. The output circuit of the mixer stage contains a tuned circuit which is resonated at the difference frequency.

The output of the mixer is fed to the first IF

amplifier and consists of a modulated intermediate frequency signal.

Q7. If the input to a mixer consists of a 795 kc RF signal and a 1250 kc oscillator signal, what major frequencies would be present in the output of the mixer?

Q8. What would the value of the difference frequency for question 7 be?

31-12. IF Amplifier

Superheterodyne receivers employ one or more IF amplifiers depending on design and quality of the receiver. Transformers are usually used for interstage coupling in the IF section. The IF circuits are permanently tuned to the difference frequency between the incoming RF signal and the local oscillator. As previously stated, all incoming signals are converted to the same frequency by the mixer stage, and the IF amplifier operates at only one frequency. The tuned circuits, therefore, are permanently adjusted for maximum gain consistent with the desired bandpass and frequency response. These stages operate as class A voltage amplifiers and practically all of the selectivity (adjacent broadcast channel and interference frequencies, not image frequency) of the superheterodyne receiver is determined by them.

The values of intermediate frequencies for broadcast receivers range from 130 kc to 485 kc, with the most popular value being around 455 kc. The choice of the IF will be discussed in Chapter 35.

The output of the final IF amplifier is fed to the detector, or demodulator.

31-13. Detector, AF Amplifier, and Reproducer

The receiver functions of detection, AF amplification, and reproduction are performed by the detector, AF amplifier, and speaker in the same manner previously described for the TRF receiver.

31-14. Characteristics of the Superheterodyne Receiver

Since the IF stages operate at a single frequency, the superheterodyne receiver may be designed to have better selectivity across the entire broadcast band and better gain per stage than the TRF receiver.

The major disadvantage of the "superheterodyne" is the reception of image frequencies. However, this disadvantage has been eliminated, to a large extent, due to the improved design of preselector circuits and the controlling of station broadcast patterns.

A6. 535 kc to 1605 kc.

A7. 795 kc, 455 kc, 1250 kc, and 2045 kc.

A8. 455 kc.

EXERCISE 31

1. Define the term "sensitivity".
2. Define the term "selectivity".
3. What is the function of the RF amplifiers and interstage coupling networks in the TRF receiver?
4. How may the sensitivity and selectivity of a TRF receiver be increased?
5. Why is a large circulating current desired in the input tank of the 1st, RF amplifier?
6. How is the bandwidth of the transmitted signal described?
7. Why cannot the bandwidth of the RF stage be less than the bandwidth of the transmitted signal?
8. What are the two functions of the detector stage?
9. Give some disadvantages of the TRF receiver (compared to the superheterodyne).
10. Explain the main difference between the TRF and the superheterodyne receiver.
11. Why must the local oscillator in a superheterodyne be tuned at the same time the preselector is tuned?
12. How does the output of the mixer differ from the input of the mixer?
13. Give some advantages of the superheterodyne (compared to the TRF receiver).
14. Explain the major disadvantage of the superheterodyne (compared to the TRF receiver).

CHAPTER 32

TRANSISTOR RF AMPLIFIERS

The function of a radio receiver is to intercept a small percentage of the radio-wave energy radiated from the antennas of transmitters and recover the original intelligence contained in it.

The transmitter of a radio station may have an output power on the order of thousands of watts, and the voltage on the transmitting antenna may be on the order of thousands of volts. However, the value of RF energy intercepted by the receiver depends on many factors such as: the distance between the transmitter and the receiver, the location of the receiver (in a valley, on a hill, behind a building, etc.), the type of receiving antenna used (directional, high gain, etc.), orientation of the receiver antenna, the power of the transmitter, and the type of terrain over which the signal is passed on its way to the receiver. Therefore, after taking all of these factors into consideration, it is not surprising to find that usually the receiving antenna only intercepts a few microwatts of power, and the voltage induced in the antenna is measured in microvolts.

In the early days of radio, receivers were rather simple devices. They performed the minimum basic requirements of a receiver, which are reception, selection, detection, and reproduction. A schematic diagram of one of these simple receivers, called CRYSTAL SETS, is shown in Figure 32-1.

The antenna performs the function of reception. Selection of the desired signal is accomplished by adjusting the variable tuning coil, L_1. The crystal detector (CR_1) rectifies the signal, and the capacitor (C_1) filters the RF component of the detected signal. The audio component of the signal is then passed on to the earphones, which perform the function of reproduction by converting the electrical signal into sound waves.

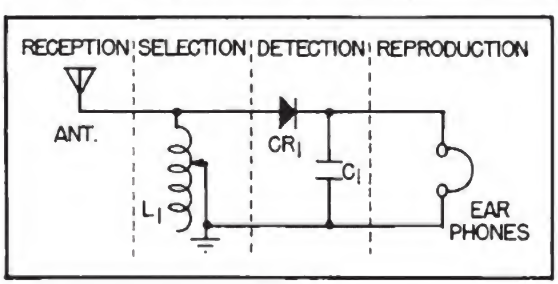

Figure 32-1 - Crystal set schematic.

It was difficult for a family to enjoy a radio broadcast when only one person at a time could use the earphones. This problem led, by popular demand, to another function being added to the receiver—audio amplification. Now the set produced enough power to operate a speaker, and everybody could hear the reproduced sound waves at the same time. Reception was still difficult, and as the broadcast spectrum began to fill up with stations, it became increasingly more difficult to separate one station from the other. It was discovered that the addition of RF amplifier stages not only increased the sensitivity of the set (allowed reception of weaker stations), but also its selectivity (allowed better discrimination between adjacent stations).

In the search for improved reception 3, 4, and sometimes 6 stages of tuned radio frequency amplification were added to the receivers. This led to increased problems in tuning and neutralization.

Many of the problems of the TRF receivers were reduced, or eliminated, by the introduction of the superheterodyne principle of reception, and today practically all of the radio receivers use this principle. The tuned RF amplifier is not necessary to the operation of a superheterodyne receiver, but one or more RF amplifiers are included in higher quality receivers where better reception is desired.

32-1. Block Diagram of Superheterodyne Receiver

Figure 32-2 shows the position of the RF amplifier in relation to the other stages of the superheterodyne receiver. The RF amplifier stage receives its input from the antenna and sends an amplified output to the mixer stage for conversion to the receiver's intermediate frequency (IF).

Besides amplification of the received RF signal, the RF amplifier stage serves other important functions. Due to its location, the RF amplifier isolates the local oscillator from the receiver's antenna. Without isolation the local oscillator frequency could be radiated from the receiver's antenna. Such radiation could interfere with the performance of other electronic

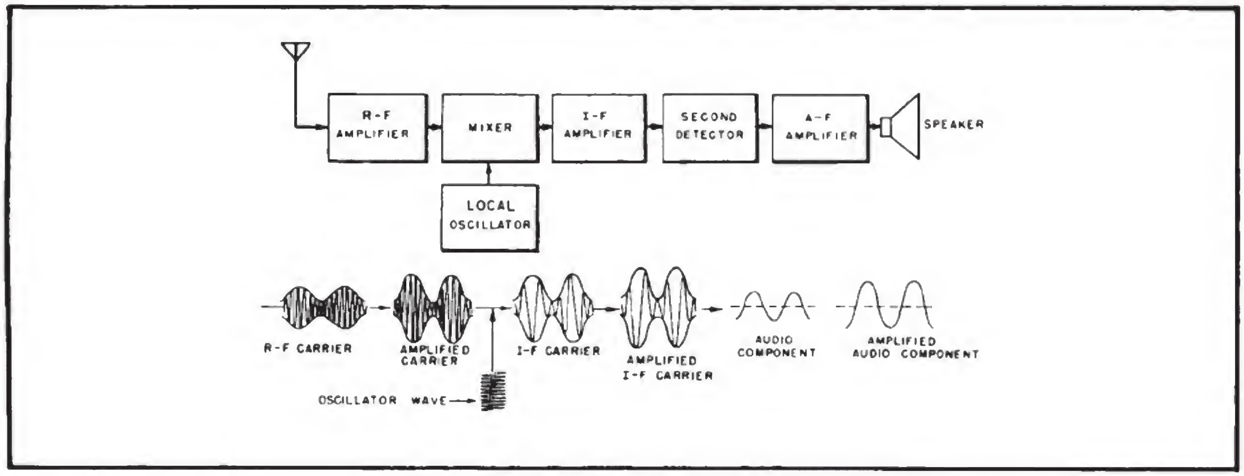

Figure 32-2 - Block diagram of superheterodyne receiver and waveforms.

equipment in the immediate vicinity of the receiver; and even more important, the radiated signal could be detected by enemy radio direction finding (RDF) equipment; thus, revealing the position of the receiver. For this reason and others, Navy superheterodyne receivers are provided with at least one RF amplifier stage. Another function of the RF amplifier is to provide increased selectivity for the receiver.

Having established a need for RF amplifiers, this chapter will provide information concerning transistor circuits used for this purpose. Included will be schematic diagrams and the information concerning the selection of the most desirable circuit configuration, bandwidth characteristics, and methods of tuning.

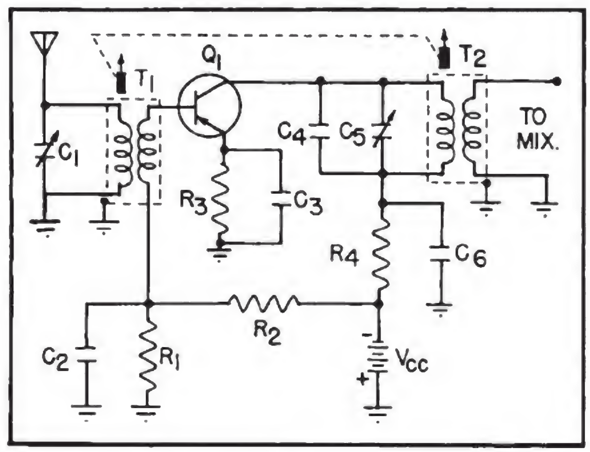

Figure 32-3 - Transistor RF amplifier.

32-2. General Operation of RF Amplifier

Before entering into a detailed analysis of RF amplifiers, the schematic diagram and general operation of a typical RF amplifier will be discussed. Figure 32-3 illustrates the schematic diagram of a transistor circuit used as an RF amplifier.

The amplifier uses a PNP transistor in the common emitter configuration. The primary of T_1 forms a tank circuit with C_1 which is part of the receiver's antenna circuit. Any broadcast signal will induce a current in the antenna; however, only currents at or near the tank's resonant frequency will cause tank oscillations to occur. The tank oscillations are coupled from primary to secondary of T_1, and applied to the base of transistor Q_1. This causes collector current to vary in accordance with the applied signal, with a resultant amplified signal developed across the collector tank circuit (C_4, C_5, and T_2 primary).

The dc operating point of the transistor is established by the voltage divider action of resistors R_1 and R_2. Stabilization and emitter bias is provided by R_3. C_3 prevents degeneration by placing the emitter at ac ground. C_2 and C_6 act as decoupling capacitors, and prevent the development of RF variations across the power supply V_{CC}. R_4 is a voltage dropping resistor used to supply the proper collector voltage to Q_1. The collector tank circuit is resonant to the same frequency as the antenna tank circuit. This provides greater selectivity for the stage. The output of the RF amplifier is transformer coupled to the mixer stage by T_2.

T_1 and T_2 are permeability, or slug-tuned, transformers whose powdered iron cores (slugs) are variable and are ganged together to provide a means of tuning the receiver through the broadcast band. When the tuning control is turned to select a broadcast signal, the inductance of T_1 and T_2 primaries will be varied, causing the respective tanks to resonate at the frequency of the desired signal. C_1 and C_5

are made variable for alignment purposes (to be explained).

Q1. When the antenna tank circuit, C_1-primary of T_1, is resonant to a specific signal frequency, does it present a maximum or a minimum impedance to that signal frequency?

32-3. Selection of Configuration

The purpose of this section is to show how the requirements for the operation of an RF amplifier have a bearing on the selection of the transistor configuration to be used.

Why is the common-base (CB) configuration better in one application and the common-emitter (CE) configuration better in another? Before the best configuration for a particular application can be chosen, there are many factors to be considered. Some of the more important of these factors are listed below:

1. The type of circuit and class of operation to be used.
2. The type (power, voltage, or current) and amount of gain required.
3. The range of operating frequencies to be encountered.
4. Impedance matching between input and output circuits.
5. Ease of neutralization for high frequency operation.
6. Amount of internally generated noise that is acceptable for a particular application.

Some of the factors involved in the choice of a configuration are solved by the requirements of an RF amplifier. One of the most basic requirements of an RF amplifier is that it will amplify the input signal voltage in a distortion free manner. The fact that the circuit is to be primarily a voltage amplifier eliminates the common-collector configuration, since this configuration exhibits a voltage gain of less than one. To amplify the signal in a distortion free manner requires the transistor to be biased for class A operation.

The type of circuit required has narrowed the selection to either the CB or CE configurations. Either can be biased class A and either will give a voltage gain.

ALPHA CUT-OFF FREQUENCY is the frequency at which the gain of a CB configuration drops to 70.7% of its gain measured at 1 kc. BETA CUT-OFF FREQUENCY is the frequency at which the gain of a CE configuration drops to 70.7% of its gain measured at 1 kc. The average transistor used as an RF amplifier, for broadcast band reception, has an alpha cut-off frequency (f_{co}) of approximately 40 Mc. The beta cut-off frequency (f_{Bco}) is determined by use of the following equation:

$$f_{Bco} = (1 - \alpha) f_{\alpha co} \qquad (32\text{-}1)$$

where: f_{Bco} = beta cut-off frequency in cps

$f_{\alpha co}$ = alpha cut-off frequency in cps

α = alpha of the transistor

Thus, a transistor with a current gain (α) of approximately 0.98 and an $f_{\alpha co}$ of 40 Mc will have a f_{Bco} of:

$$f_{Bco} = (1 - \alpha) f_{\alpha co}$$

$$f_{Bco} = (1 - 0.98)\ 40 \times 10^6$$

$$f_{Bco} = 0.02 \times 40 \times 10^6$$

$$f_{Bco} = 800 \text{ kc}$$

From the above it can be seen that the gain of a transistor used in the CE configuration will decrease to the 70.7% point at a much lower frequency than the same transistor used in the CB configuration. A comparison of gain versus frequency for the two configurations is shown in Figure 32-4. The curves are for a typical transistor used in RF amplifier applications in AM broadcast band receivers. The curves are plotted with the output terminals of the transistor ac shorted to remove load influences.

The transistor is a 2N1631, germanium PNP type. The 2N1631 has an alpha of 0.987, an alpha cut-off frequency of 45 Mc, and a beta of 76. The beta cut-off frequency by equation (32-1) is:

$$f_{Bco} = (1 - \alpha) f_{\alpha co} \qquad (32\text{-}1)$$

$$f_{Bco} = (1 - 0.987)\ 45 \times 10^6$$

$$f_{Bco} = 585 \text{ kc}$$

The range of frequencies covered by the broadcast band has been superimposed on the gain curves (Figure 32-4) for a comparison of the gain of the two configurations within the desired operating range. The gain of the CB configuration is low, but relatively constant within the broadcast band. Whereas, the gain of the CE configuration is high and varies considerably within the range of operating frequencies.

A1. Maximum. Because the antenna tank in this particular circuit, is a parallel resonant circuit.

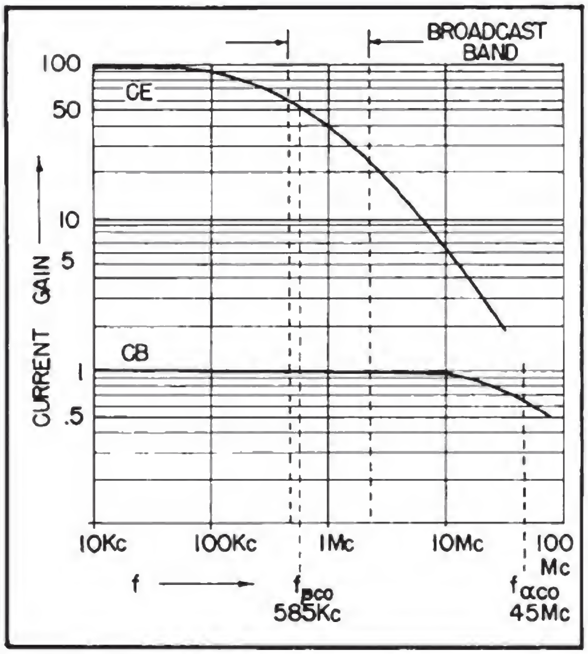

Figure 32-4 - CE and CB configuration gain versus frequency for 2N1631.

If the amplifier were to be used in a wide band circuit where gain was not of primary importance, the CB configuration would be best suited. On the other hand, if the application required a narrow band amplifier and gain was of prime importance, the CE configuration would be more satisfactory.

The required bandwidth of a single stage RF amplifier for a broadcast band superheterodyne is very small compared to the operating frequency of the circuit. For example, assume a bandwidth of 40 kc. The ratio of operating frequency to bandwidth even at the low end of the broadcast spectrum is:

$$\frac{550 \text{ kc}}{40 \text{ kc}}$$

or approximately 13.7 to 1. While at the high end of the spectrum the ratio is:

$$\frac{1600 \text{ kc}}{40 \text{ kc}}$$

or approximately 40 to 1. Thus, in this application, a bandwidth of 40 kc is considered a narrow band amplifier. However, if the same 40 kc bandwidth were used with an operating frequency of 80 kc, the ratio of operating frequency to bandwidth would be only:

$$\frac{80 \text{ kc}}{40 \text{ kc}}$$

or 2 to 1. In this application a bandwidth of 40 kc would be considered a wide band.

Since the RF amplifier is a narrow band amplifier, and the bandwidth is such a small portion of the broadcast spectrum, the variation of gain of the CE configuration within the bandwidth would be of little consequence. If it were absolutely necessary to have the gain of the CE and the flat response of the CB (and additional cost did not matter), a transistor having an f_{Bco} above the broadcast band could be chosen. Two such transistors are the 2N700 with an f_{Bco} of 7 Mc, and the 2N509 with an f_{Bco} of 10.5 Mc.

The ratio of output to input resistance is smaller (values closer together) in the CE than in the CB. Thus, the problem of matching the output impedance of one stage to the input impedance of the following stage is greatly simplified when CE configurations are used.

At broadcast band frequencies the neutralization of either configuration (if required) is a relatively simple accomplishment. The transistors used as RF amplifiers in AM receivers typically yield 25-30 db gain unneutralized and 28-33 db gain when neutralized. However, at higher frequencies neutralization becomes increasingly more complex. At frequencies approaching the alpha cut-off of the transistor, neutralization of the CB configuration is much simpler than for the CE (due to the superior isolation between the input and output circuits of the CB). At high frequencies the use of the CB configuration results in a more stable circuit. The operation of the CE configuration becomes unstable and unreliable at very high frequencies.

The internally generated noise of the transistor (discussed in a later section) is about the same for either the CB or CE configurations. The selection of a configuration is not appreciably affected by internal noise considerations.

In summary it can be said that in narrow band high gain applications and where variation in stage gain within the bandwidth is not critical, or where relative simplicity in design is desired, the CE configuration may be expected to predominate. In very wide band applications where constant gain within the bandwidth is important, stage gain is not a critical factor; or where very high frequency operation with relatively narrow bandwidths is employed, the CB configuration may be expected to predominate.

Q2. If the alpha cut-off of a transistor is 100 Mc and the alpha is 0.99, what is the cut-off frequency when this transistor is used in the CE configuration?

Q3. Would a two stage CE or a two stage CB amplifier be simpler to design? Why?

32-4. Input Circuit

Small currents are induced in receiver antennas when the antenna intercepts a signal in the form of electromagnetic energy. There are many frequencies present in the antenna, including those within the broadcast band (535 kc to 1605 kc). Equal amplification (at the same time) of all signals present in the antenna would result in excessive interference. It is desirable to select a specific station's signal and reject all other signals. However, for reasons yet to be discussed, the bandwidth of the RF amplifier must be wider than the exact bandwidth of the desired station signal. Therefore, adjacent station signals and interference frequencies above and below the desired station frequency will receive some amplification.

The ability of a tank circuit to select a band of frequencies is a function of its Q. The higher the Q, the smaller is the bandwidth. As will be explained in Chapter 35, the selectivity of a superheterodyne receiver is determined primarily by the IF circuits. Since the receiver does not depend exclusively on the input circuit for selectivity, the Q of the input circuit need only be high enough to provide good image rejection. Figure 32-5 illustrates the relationship of the image frequency to the response curve of the tuned input circuit (for a specific value of f_o). The curve is arbitrarily chosen, its shape is representative of the curve for tuned circuits used in RF amplifiers. Figure 32-5 shows that (providing the Q of the input circuit is sufficiently high and the amplitude of the f_o is not many times smaller than the image frequency amplitude) the possibility of the image frequency receiving enough amplification to pass through the receiver is very small.

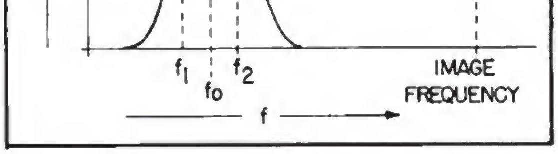

Figure 32-5 - Image frequency and input circuit response curve.

Q4. Would the image frequency be closer to the f_o in a receiver employing a 262 kc IF or one employing a 455 kc IF?

Q5. To prevent image frequency reception, would the selectivity of the input circuit have to be better in a receiver employing a 465 kc IF or in one with a 175 kc IF?

BANDWIDTH CHARACTERISTICS

The main reasons why an RF amplifier, or preselector, is used in a receiver is to provide:

1. Amplification.
2. Isolation.
3. Image rejection, or increased selectivity.
4. Improved signal to noise ratio (to be discussed).

Amplification and isolation are provided by the action of the transistor used in the preselector. On the other hand, selectivity and image rejection (and to a certain extent, signal-to-noise ratio) are mainly a function of the bandwidth of the stage. The bandwidth is determined primarily by the tuned circuits and coupling networks of the stage. The trainee desiring to refresh his memory on the effect of Q on bandwidth and selectivity is directed to review sections 11-47, 11-48, 12-17, and 12-18.

32-5. RF Amplifiers in Cascade

In some cases a single stage of RF amplification is insufficient to provide the required signal input to the mixer stage. Under these conditions two or more RF stages may be connected in cascade. When RF circuits (transformer coupled) are connected in cascade, the resultant bandwidth is smaller than that of a single stage. The reason for this can be explained in the following manner: In order to simplify the explanation, the amplifiers associated with the RF stages will be assumed to have an amplification factor of unity. Figure 32-6 is a block diagram of two cascaded RF stages. Notice, while there are only two amplifiers, three tuned circuits are employed. The capacitors of the tuned circuits are ganged to permit adjustment of the three circuits to the same resonant frequency. Further assumptions to be made in this discussion are: the three tuned circuits have identical Q's, and there is

A2. f_{Bco} equals 1 Mc.

A3. CE. The ratio of output to input impedance is smaller for the CE configuration.

A4. Image frequency would be closer to f_o in the receiver with the 262 kc IF.

A5. Better selectivity would be required in the receiver with the 175 kc IF.

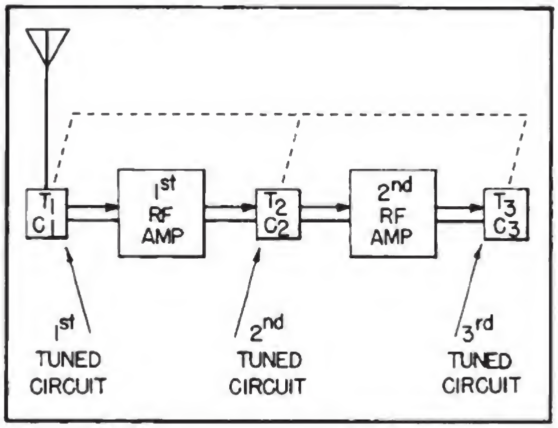

Figure 32-6 - Block diagram of cascaded RF amplifiers.

no resonant voltage rise in the secondaries of the transformers T1, T2, and T3.

Three frequencies of equal amplitude are applied to the input of the two stages; the resonant frequency f_o and two frequencies, f_y and f_x, equidistant above and below the resonant frequency.

The response curve for each INDIVIDUAL stage will be identical, and will appear similar to the curve in Figure 32-5. However, the OVERALL response curve of the three circuits will differ from the individual curves.

Figure 32-7 shows the overall response curves produced by cascading tuned circuits.

Curve 1 represents the response curve of the first tuned circuit. Curve 2 is the OVERALL response curve of the first RF amplifier stage (including the first and second tuned circuit). Curve 3 is the OVERALL response curve of the two RF amplifiers in cascade (including the three tuned circuits).

The maximum voltage is developed across the tuned circuits at the resonant frequency, and is taken as a 100% for purposes of references. The frequencies, f_y and f_x, coincide with the half-power points of the first tuned circuit.

If the maximum voltage developed across the first tuned circuit is 10 volts, the voltage de-

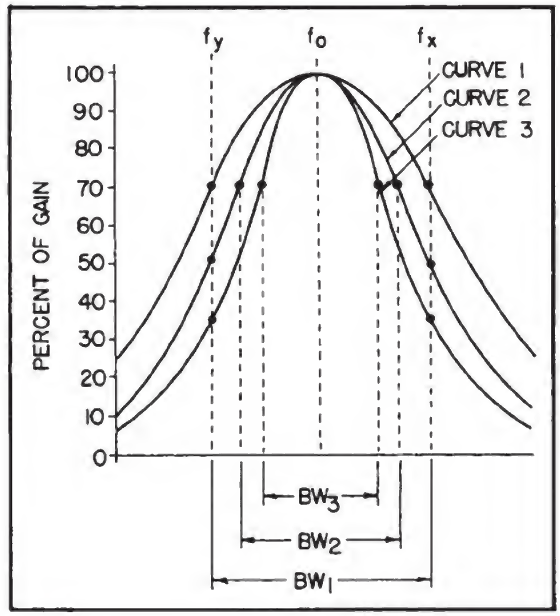

Figure 32-7 - Response curves for cascaded-tuned circuits.

veloped by f_y and f_x at the half-power points will be equal to:

$$0.707 \times 10 = 7.07 \text{ volts}$$

The bandwidth of the first tuned circuit is shown as BW_1 in Figure 32-7.

The three frequencies (f_o, f_y, f_x) would normally be amplified by the first RF amplifier and applied to the output tank, which is the second tuned circuit; but in this explanation they are passed through the amplifier with no amplification.

Since the tuned circuits are identical the resonant frequency will develop 10 volts across the second tuned circuit also:

$$100\% \times 10 \text{ volts} = 1 \times 10 = 10 \text{ volts}$$

The frequencies f_y and f_x (whose amplitude was reduced to 7.07 volts by the 1st tuned circuit) now appear at the half-power points of the 2nd tuned circuit. Thus, the voltage developed by f_y and f_x across the 2nd tuned circuit will be:

$$0.707 \times 7.07 = 5 \text{ volts}$$

The overall response curve (curve 2, Figure 32-7) for the 1st and 2nd tuned circuits in cascade shows the gain at f_y and f_x to be down to 50% of maximum. The separation between the half-power points of curve 2 is less than for

curve 1. It can be seen that the bandwidth (BW_2) for two tuned circuits is less than for a single tuned circuit.

The three frequencies are passed through the 2nd RF amplifier and applied to the 3rd tuned circuit. The original amplitude of f_o is unchanged, and will develop 10 volts across the 3rd tuned circuit. On the other hand, the amplitude of f_y and f_x has been reduced to 5 volts by the 2nd tuned circuit. Since f_y and f_x appear at the half-power point of the individual response curve for the 3rd tuned circuit they will develop a voltage across this circuit of:

$$0.707 \times 5 = 3.54 \text{ volts}$$

Curve 3 (Figure 32-7) indicates that passing f_y and f_x through three tuned circuits has attenuated their amplitudes until they are now only 35.4% of the maximum value. The 70.7% points of curve 3 have moved even closer to f_o making the bandwidth (BW_3) still smaller. It can be seen from the preceding explanation that with the addition of each stage containing a tuned circuit, the bandwidth will decrease. The above also explains the necessity of having the bandwidth of each individual circuit larger than the required overall bandwidth of the receiver. Figure 32-7 also indicates that the addition of stages containing tuned circuits has increased the slope of the overall response curve; thereby, increasing the selectivity of the receiver.

Q6. In what manner is the fidelity (ability of a receiver to reproduce all of the sideband frequencies associated with the original signal) of a receiver affected by the addition of too many tuned stages?

32-6. Noise

It has been stated the sensitivity rating of a receiver indicates the minimum signal required to produce a specified output. However, various types of noise will be present in the receiver input along with the desired signal. Since noise voltages will be amplified in the same manner as the desired signal, the ratio of the noise voltage will be the limiting factor on the maximum receiver sensitivity.

There are two kinds of noise which must be considered in the operation of a receiver. They are: noise which is generated externally to the receiver, and noise which is generated within the receiver itself.

External noise may be classified in two general categories, natural and man-made. The most common natural noise is static, caused by electrical discharges and electrical storms. Man-made noise is the result of any device which produces an electrical arc, such as an internal combustion engine, neon signs, electric motors, etc.

Internal noise may also be classified in two general categories. The first category consists of THERMAL NOISE, and DIFFUSION RECOMBINATION NOISE (sometimes called SHOT NOISE). The second category consists of SURFACE NOISE and LEAKAGE NOISE. The two noises in the second category are usually grouped under the common heading of SEMICONDUCTOR NOISE. Internal noise sources will be discussed further in the topic on transistor noises (32-8).

Noise is the limiting factor for receiver sensitivity because intelligible reproduction of the signal is no longer possible when the signal strength decreases to the point where it becomes equal to the noise voltage (external plus internal noise).

32-7. Signal-to-Noise Ratio

Noise voltages having the same frequencies as the desired signal will receive a proportionate amount of amplification. Thus, the amplitude of the voltage induced in the antenna must be sufficiently large in relation to the amplitude of the noise voltages to overshadow their effects. The ratio of the desired signal amplitude to the noise voltage amplitude is known as the SIGNAL-TO-NOISE RATIO. The effect of noise voltages, whose frequencies do not lie within the bandwidth of the desired signal, can be minimized by decreasing the bandwidth (increase selectivity) of the input circuits as much as possible WITHOUT decreasing the overall bandwidth below the required amount. Therefore, decreasing the bandwidth results in an increased signal-to-noise ratio.

32-8. Transistor Noises

Internal noise need only be considered when the magnitude of the signal voltage is of the same order. Internally generated noise is always present; however, when the magnitude of the signal voltage is many times larger than the noise voltage, the noise may be neglected. Because of this, internal noise is only important in the input circuits of a receiver. Due to the complexities involved in the study of noise generation within the transistor, only the general types and their effects will be discussed in this topic.

The main difference in the concept of noise voltages and currents in relation to the normal concept of voltages and currents is that noise is completely erratic and unpredictable. When a situation such as this exists, it is described as RANDOM. One important term used in the description of noise effects is "POWER SPECTRUM". The term "power spectrum" refers to the amount of power in a narrow band of fre-

774-639 O - 65 - 6

A6. The addition of too many tuned stages may increase the selectivity to the point where the bandwidth is no longer wide enough to accomodate all of the sideband frequencies, resulting in the distortion of the signal.

---

quencies and the variation of this power as the narrow band of frequencies is varied over a whole range of frequencies.

In this discussion there are two types of power spectrums that will be considered. The first is a "flat" power spectrum. In a flat power spectrum the amount of power in the narrow band of frequencies does not change as it is varied from one end of a frequency range to the other. The second type is a power spectrum that varies with frequency. In this type, the power in the narrow band of frequencies varies inversely with frequency (1/f).

The most basic type of noise in transistors, and associated circuitry, is noise produced by the random motion of current carriers. This random motion of current carriers occurs within all parts of the circuit—the transistor as well as every conductor, resistive component, etc. Random carrier motion is caused by thermal agitation; therefore, the resulting noise is called thermal noise. Thermal agitation also causes a continual random generation of electron-hole pairs (within the semiconductor material only) which is accompanied by a continual random recombination of free electrons and holes. The noise generated by the random recombination and collisions of mobile carriers with fixed atoms, is called diffusion recombination noise. Both thermal noise and diffusion recombination noise exhibit a flat power spectrum. In general, noises with a flat power spectrum are termed WHITE NOISE.

The explanation of the physical causes of surface noise and leakage noise are rather involved and no attempt will be made to describe them here. These noises, however, are peculiar to semiconductors and so their effects are of interest in the study of transistor circuits. Semiconductor noise exhibits a power spectrum that varies inversely with frequency, for this reason it is sometimes called 1/f noise.

A plot of noise power versus frequency for a typical junction transistor shows the frequency range as being separated into two regions. At low frequencies there is a region where the circuit noise is caused by semiconductor noise. As frequency is increased the 1/f noise decreases and eventually a point is reached where the 1/f noise becomes less than the white noise. For a typical junction transistor this crossover point usually occurs somewhere in the audio frequency range (1 kc to 10 kc). From the crossover point to a point beyond the alpha cut-off frequency, the circuit noise is caused by white noise. Therefore, it remains relatively constant. At frequencies beyond the alpha cut-off of the transistor the circuit noise begins to increase due to the reduction in circuit gain.

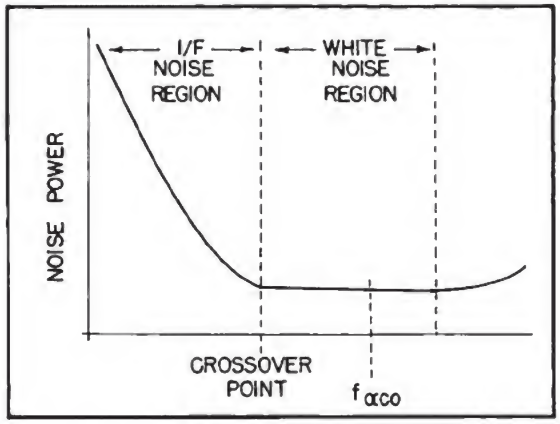

Figure 32-8 - Transistor noise power versus frequency.

Figure 32-8 shows a plot of noise power versus frequency and the important noise regions.

In a well designed circuit, with a reasonably good transistor, the internal noise will be a very small portion of the total noise. Thus, the external noise, arriving with the signal, will be the determining factor in the signal-to-noise ratio.

Q7. In what manner does decreasing the bandwidth increase the signal-to-noise ratio?

Q8. What type of noise is predominant in a circuit if the noise power within the bandwidth varies as the operating frequency is changed?

32-9. Two Stage RF Amplifier

Figure 32-9 illustrates the schematic of a two stage RF amplifier connected in cascade. Ganged capacitive tuning is used in the tuned circuits to select the resonant frequency. The common emitter configuration is used in order to take advantage of the greater power gain and the simpler biasing arrangements. Biasing for the operating point of Q_1 and Q_2 is determined by the voltage divider method, using R_1, R_2, R_7, and R_8. Emitter bias and temperature stabilization for the two transistors are provided by R_3 and R_9. C_4 and C_8 prevent degeneration across the stabilizing resistors. C_{14} and C_{15} are neutralizing capacitors. R_6 and R_{10} determine the amount of collector voltage

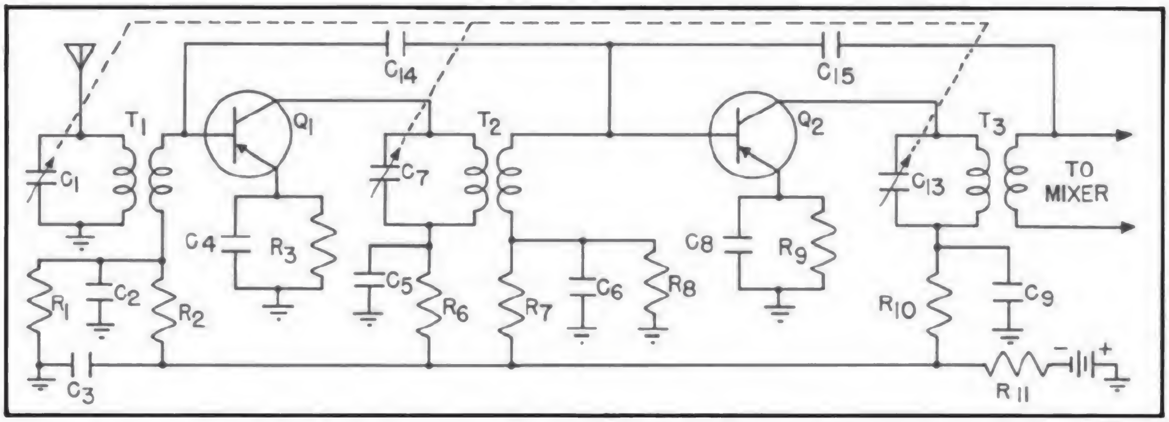

Figure 32-9 - Two stage RF amplifier.

for Q_1 and Q_2 respectively. Capacitors C_2, C_3, C_5, C_6, and C_9 bypass the RF currents around the biasing networks and the power supply.

TUNING

It has already been pointed out that one of the prime functions of the first tuned RF circuit, in a receiver, is that of selecting the desired signal frequency from among the many frequencies available. The process of adjusting the inductance (L) or the capacitance (C) of a tuned circuit in order to selecta specific station frequency is called TUNING.

32-10. LC Product

The resonant frequency of a tuned circuit is determined by the product of the capacitance and inductance, called the LC PRODUCT. For every single resonant frequency there is a single value of the LC product. If two resonant circuits have the same resonant frequency they must have the same LC product.

The LC product for a given frequency may be determined by use of equation (32-2).

$$LC = \frac{1}{\omega^2} \qquad (32\text{-}2)$$

Where: LC = the product of inductance in henries and capacitance in farads

$$\omega = 2\pi f$$

NOTE: Equation (32-2) results from transposition of equation (11-16).

$$f_o = \frac{1}{2\pi\sqrt{LC}} \qquad (11\text{-}16)$$

For a station frequency of 540 kc the LC product will be:

$$LC = \frac{1}{\omega^2} \qquad (32\text{-}2)$$

$$LC = \frac{1}{(6.28 \times 540 \times 10^3)^2}$$

$$LC = 0.0871 \times 10^{-12}$$

Knowing the LC product and the value of either the inductance or capacitance will permit the determination of the value of the unknown component. For instance, assume the inductance used in the tuned input circuit is approximately 256 micro-henries. In order to find the value of capacitance which must be used to make the circuit resonant at 540 kc, divide the inductance into the LC product:

$$C = \frac{LC}{L}$$

$$C = \frac{0.0871 \times 10^{-12}}{256 \times 10^{-6}}$$

$$C = 340 \text{ pf}$$

A receiver may be tuned by varying either the inductive or the capacitive component of the tuned circuit. Assuming that the inductance is a fixed value, then the capacitance must be varied in order to tune the receiver across the broadcast band. It has already been determined that, in order to tune the receiver to the lowest frequency in the receiver's range, the variable

A7. Noise outside the bandwidth receives less amplification while the amplification at the signal frequency is maintained the same.

A8. Semiconductor noise.

---

tuning capacitor must equal 340 pf at one of its limits. The other limit of the variable capacitor may be found by determining the LC product for the highest frequency in the receiver's range and dividing by the inductance, as follows:

High frequency is approximately 1600 kc. Thus:

$$LC = \frac{1}{\omega^2} \qquad (32\text{-}2)$$

$$LC = \frac{1}{(6.28 \times 16 \times 10^5)^2}$$

$$LC = \frac{1}{10100 \times 10^{10}}$$

$$LC = 0.00992 \times 10^{-12}$$

Determine capacitance by dividing the inductance into the LC product:

$$C = \frac{LC}{L}$$

$$C = \frac{0.00992 \times 10^{-12}}{256 \times 10^{-6}}$$

$$C = 38.7 \text{ pf}$$

From the above it can be seen that the first tuned circuit of a radio receiver, designed to cover the broadcast band, could be constructed using a 256 microhenry coil and a variable capacitor having a range of 38.7 pf to 340 pf. A practical tuning capacitor would probably range from 30 to 350 pf.

As shown previously, a receiver has more than one tuned circuit. If all the tuned circuits are tuned by varying the capacitance, then an air dielectric variable capacitor having two or more sections is used. All sections are mechanically connected to the same shaft, but electrically isolated from each other. Each section is used for a separate tuned circuit. When the shaft is moved, the capacitance of all tuned circuits will vary by the same amount. Figure 32-10

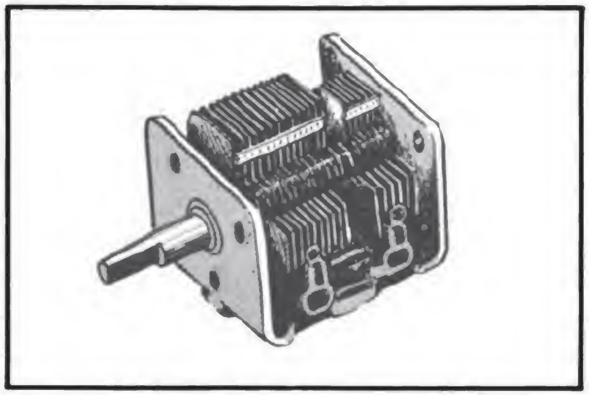

Figure 32-10 - Ganged air dielectric variable capacitor.

illustrates a ganged air dielectric variable capacitor.

In circuits that are tuned inductively, the cores of the inductors are mechanically connected together to permit changing the inductance of all tuned circuits simultaneously.

Q9. What is meant by "tuning" a receiver?

Q10. The LC product of a tuned circuit is 0.045 times 10^{-12}. The variable capacitor is adjusted to 150 pf. What is the value of the fixed inductance used in the circuit?

32-11. Trimmers and Padders

It was stated that a ganged variable capacitor will vary the capacitance of all the tuned circuits an equal amount. However, due to slight mechanical differences (manufacturing tolerances) in the capacitors, coils, and variations in wiring from one receiver to another, there will be minor variations in electrical adjustment between individual sections. In order to obtain maximum performance all tuning circuits must "track" together. Tracking means that the resonant frequency for each tuning circuit will be the same for all positions of the tuning shaft. To compensate for differences in capacitance, between the ganged sections, TRIMMER capacitors and PADDER capacitors are used.

Trimmers and padders are small variable capacitors consisting of two plates separated by a sheet of mica. The trimmer is connected in parallel with the tuning capacitor and the padder is connected in series with it. Figure 32-11 illustrates the connection of a trimmer and padder capacitor. At the high frequency end of the tuning range, the tuning capacitor is set for MINIMUM capacitance. The parallel trimmer has about the same order of magnitude as this minimum value, and its adjustment determines the

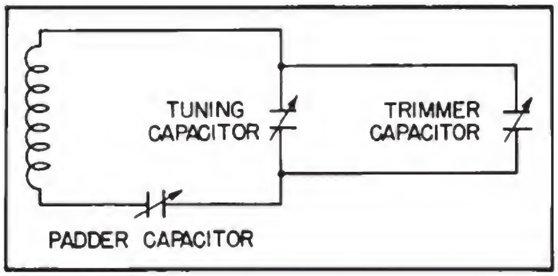

Figure 32-11 - Trimmer and padder connections.

proper resonant frequency at the high end of the frequency range.

At the low frequency end of the tuning range, the tuning capacitor is set near maximum capacitance, and therefore the small parallel trimmer is negligible in comparison with it. However, the series padder is comparable in magnitude to the value of the main tuning capacitor, and adjustment of the padder will affect the resonant frequency at the low end of the tuning range. Since the total capacitance of two series capacitors is influenced chiefly by the smaller of the two, the effect of the series padder on the total tuning capacitance at the high end of the tuning range is negligible. By proper adjustment of the trimmer and padder capacitors resonable tracking accuracy can be obtained throughout the frequency range.

Q11. How is a trimmer capacitor connected in a circuit?

Q12. How is a padder capacitor connected in a circuit?

A9. Varying either the capacitance or inductance of a tuned circuit in order to select a specific station frequency.

A10. 300 microhenries.

A11. In parallel with the tuning capacitor.

A12. In series with the tuning capacitor.

EXERCISE 32

1. What is the purpose of the RF amplifier?
2. What is the function of the first tuned circuit?
3. Explain why the common-emitter configuration is desired as an RF amplifier in the broadcast band.
4. If a transistor has a beta of 100, what is its alpha?
5. Why is neutralization of the CE more difficult than the CB at frequencies approaching alpha cut-off frequency?
6. Explain the two broad categories of receiver noise.
7. Explain the two general categories of noise generated by a transistor.
8. What is meant by receiver sensitivity?
9. Explain how receiver selectivity effects fidelity.
10. Define the term bandwidth.
11. Explain how the cascading of tuned RF amplifiers effects overall bandwidth. Why?
12. What are some advantages of cascaded RF amplifiers as compared to a single stage?
13. What is meant by tracking?
14. At which end of the band is the trimmer capacitor adjusted. Explain why.
15. At which end of the band is the padder capacitor adjusted? Explain why.

CHAPTER 33

TRANSISTOR OSCILLATORS

The rapid alternating motion (oscillation) of electrons in a conductor results in radiation of electromagnetic waves, and this phenomenon forms the basis of all radio communication. The term oscillate is defined as "to swing or move back and forth." An OSCILLATOR is defined as "a device which oscillates or produces oscillations." The purpose of an electronic oscillator is to produce the rapid back and forth, or alternating, motions of electrons (called alternating current) from a direct current supply.

Previous chapters have discussed the electron tube oscillators and their role in relation to transmitter operation. The purpose of this chapter is to discuss transistor oscillators and their relation to the operation of receivers. Although different in overall appearance and operation, many of the fundamental actions of a specific type of transistor oscillator are similar to those of its electron tube counterpart. Therefore, where applicable, the actions of comparable types of transistor and electron tube oscillators will be compared.

The configurations and theory of operation of the transistor Hartley and Colpitts type oscillators will be discussed. The various transistor configurations and their characteristics will be discussed in relation to oscillator operation. The factors affecting frequency and amplitude stability of transistor oscillators will be mentioned. Finally, factors involved in local oscillator tuning and methods of coupling output signals from transistor oscillators will be considered.

Many of the fundamental actions, pertaining to oscillators, referred to in this chapter will be based on information contained in chapter 24. Therefore, only brief reviews of pertinent information previously covered will be included in this chapter. The trainee desiring to refresh his memory on vague points is directed to review the applicable areas of chapter 24.

33-1. Receiver Oscillator

The oscillator in a receiver is usually referred to as a LOCAL OSCILLATOR. The block diagram of a superhet receiver (chapter 32) shows that the output of the local oscillator is one of the inputs applied to the mixer stage.

In a typical superheterodyne receiver the local oscillator (LO) generates a sine-wave of constant amplitude and frequency (when tuned to a specific station). The sine wave output signal from the local oscillator and the modulated radio frequency output signal from the RF amplifier are both applied to the mixer stage. In the mixer stage, heterodyning occurs between the oscillator signal and the output signal from the RF amplifier. As a result of this heterodyne action, a new and lower intermediate frequency (IF) is generated.

It was shown that a basic oscillator can be broken down into three main sections: a frequency determining device, an amplifier, and a feedback circuit. The frequency determining device in a superheterodyne receiver oscillator is usually an LC tank circuit. While the tank circuit is NORMALLY found in the input circuit of the oscillator (both electron tube and transistor) it should not be considered out of the ordinary if it appears in the output circuit of a transistor oscillator. The differences in magnitude of plate and collector currents and shunting impedances are partly responsible for this. In both types of circuits (solid state and electron tube) oscillations take place in the tuned circuit. Both the electron tube and the transistor function primarily as an electrical valve that amplifies, and automatically delivers to the input circuit the proper amount of energy to sustain oscillations. In both tube and transistor oscillators the feedback circuit couples energy of the proper amount and phase from the output to the input circuit in order to sustain oscillations.

SERIES-FED HARTLEY OSCILLATOR

33-2. Review of Electron Tube Hartley Oscillator

Before entering into a discussion of the transistor Hartley a brief review of the electron tube series-fed Hartley, shown in Figure 33-1, will prove advantageous.

Capacitor C_2 and the radio frequency choke (RFC) provide decoupling between the oscillator and the power supply. Although not required for operation of the oscillator itself C_2 and RFC improve the stability of the oscillator by minimizing variations in plate voltage and current. In this respect they act much the same as the "L" type filter discussed in the chapter on power

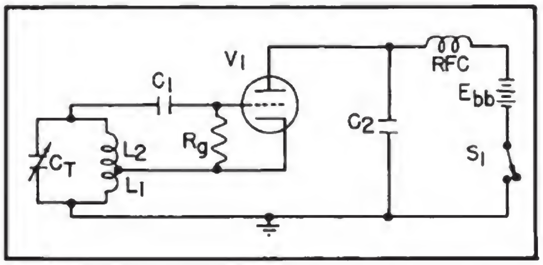

Figure 33-1 - Electron tube series-fed Hartley.

supplies. Any tendency for the plate voltage to change as a result of variations in the plate current will be suppressed. The choke offers a high resistance to changes in current, and the capacitor presents a low reactance to changes in plate voltage. When properly designed, the decoupling network greatly minimizes the effect of plate voltage and power supply current changes on the operation of the oscillator.

A close examination of Figure 33-1 shows that both the direct and alternating components of plate current flow through L_1. Since the dc plate current flows through a portion of the tank circuit the oscillator is series-fed.

Capacitor C_1 in conjunction with grid resistor R_g provides grid leak bias, the value of which determines the operating point of the tube. The tank components are L_1, L_2, and C_T.

Consider that the filaments of V_1 (Figure 33-1) are at operating temperature but switch S_1 is open. When S_1 is closed the following events occur.

1. The tube has no initial bias, therefore, the tube will begin to conduct causing current flow from ground, through L_1, through V_1, through RFC, and back to the power supply. At the same time capacitor C_2 is assuming a charge with a negative potential on the bottom plate and positive on the top plate.

2. The current flow through the feedback component L_1 induces a voltage in L_2. The induced voltage in L_2 causes a circulating current to flow in the tank which begins to charge tank capacitor C_T with a negative potential on the bottom plate and positive on the top plate.

3. The voltage across the tank is essentially the charge across C_T. Therefore, grid current will be drawn to charge capacitor C_1 as the top of the tank assumes a positive potential.

4. The regenerative cycle of increasing grid voltage causing increased plate current will continue until limited by the non-linear characteristics of the tube.

5. As the rate of change of current through L_1 decreases the induced voltage of L_2 decreases. When the value of induced voltage falls below the value of charge on capacitor C_T, the tank capacitor will begin discharging.

6. The reduction of magnitude of the positive potential at the top of the tank will cause capacitor C_1 to attempt to discharge through the large value of grid resistor R_g. Thus, applying a negative potential to the grid of V_1.

7. At the time when C_T is completely discharged circulating tank current will be maximum and the tank voltage will have completed a half cycle of operation. The magnetic fields of L_1 and L_2 will now collapse, maintaining tank current flow in the same direction and subsequently charging C_T with the opposite polarity (negative on the top plate and positive on the bottom plate).

8. At the instant the magnetic fields are completely collapsed the tank current will be zero and the voltage across C_T will be maximum. The tank circuit has now completed three fourths of a cycle.

9. Tank capacitor C_T will begin discharging, causing tank current to reverse direction. The discharge of C_T will establish a magnetic field around L_2 and L_1. When C_T is completely discharged, tank voltage will be zero, an entire cycle of oscillator operation has been completed, and the circuit is ready to begin the next cycle.

Q1. If C_2 in Figure 33-1 were to become open would the series-fed Hartley cease to operate?

33-3. Series-fed Transistor Hartley Oscillator

The circuit in Figure 33-2 shows a series-fed Hartley oscillator using an NPN type transistor. Notice that in many respects the transistor circuit is similar to the electron tube circuit. In fact, with minor limitations the transistor oscillator compares very favorably with its electron tube counterpart.

Like the vacuum tube oscillator, the transistor oscillator can be operated class A, B, or C. Where stability and pureness of waveform

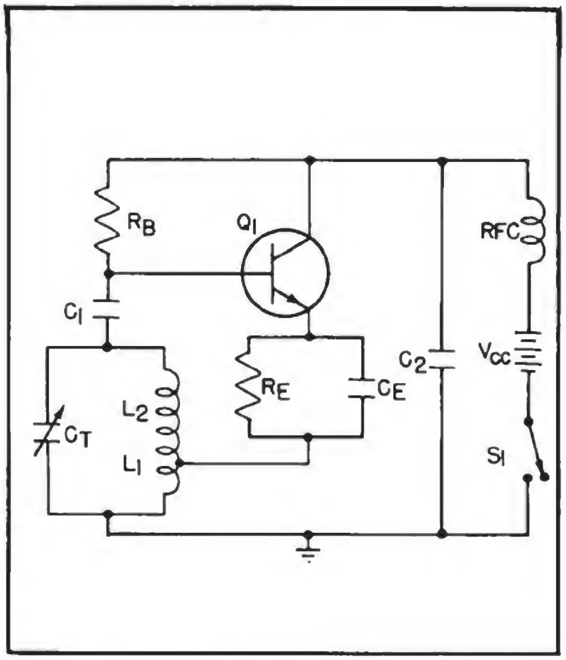

Figure 33-2 - Transistor Hartley oscillator.

(low harmonic content) are major considerations, the oscillator would be operated class A. If efficiency is the primary concern, class C operation is chosen. Inasmuch as the local oscillator in a radio receiver is not required to supply a large amount of power to the mixer, efficiency is of minor importance.

The purpose of the component parts of a transistor oscillator should become somewhat apparent upon examination of the schematic diagram in Figure 33-2. The sine wave is generated in the tank circuit consisting of C_t, L_1 and L_2. C_t is made variable so that the oscillator can be tuned to the desired frequency. The RFC and capacitor C_2 are for the purpose of decoupling. To prevent thermal runaway resistor R_E is connected in series with the emitter. Degeneration, which would otherwise occur, is minimized by connecting capacitor C_E in parallel with stabilizing resistor R_E.

In order for oscillations to begin, a small amount of initial forward bias must be applied to the emitter-base junction. This bias establishes a flow of current through the transistor when the circuit is first energized, and insures that the oscillator will be self-starting.

The magnitude of the bias current for the circuit in Figure 33-2 is determined by the values of R_B, emitter-to-base resistance, R_E, V_{CC}, and to some extent the small dc resistance contained in L_1 and RFC. The choice of values used for R_B and R_E represents a compromise. This compromise arises in the following manner. Obtaining a high Q tank circuit in an electron tube oscillator is not too difficult because there is very little shunting of the tank by parallel components (R_g, etc., are rather large values). On the other hand, the components shunting the tank circuit in a transistor oscillator are comparitively low in value. At the operating frequency capacitors C_1, C_2, and C_E have very low reactances, effectively making them short circuits for ac. Consider the circuit in Figure 33-2. The low reactance of C_1 effectively connects one end of R_B to the top of the tank while the low reactance of C_2 effectively connects the other end of R_B to the bottom of the tank (with respect to ac). Therefore, if too small a value is chosen for R_B it will swamp the tank circuit, lowering the Q to the point where oscillations cannot be sustained. If, on the other hand, R_B is made very large to reduce its shunting effect, insufficient bias may result.

Consider also that the low reactance of C_1 effectively connects the base of Q_1 to the top of the tank and the low reactance of C_E effectively connects the emitter to the tap on the tank coil. This causes the relatively low input impedance of the transistor to shunt a portion of the tank circuit. From this it becomes apparent that, while C_1 is necessary to prevent the low dc resistance of L_2 from placing a dc short across the series combination of the emitter-base junction and the stabilizing resistor, C_1 must also have sufficient reactance to provide a degree of ac isolation between the tank circuit and the shunting effects of R_B and the low input impedance of Q_1.

Due to the large variation in component values, from one comparable circuit to another, no attempt will be made to assign typical values to the components in the transistor oscillators. Although different values of a given component (within allowable limits) will alter its operation to a certain extent, the basic action of individual components in comparable circuits will remain relatively unchanged.

A complete cycle of operation for the series-fed Hartley transistor oscillator (Figure 33-2) will now be discussed in the following sections.

Q2. Starting at the negative terminal of V_{CC} give the path of bias current in Figure 33-2.

33-4. Voltage Distribution

At the instant switch S_1 is closed all the capacitors in the circuit, having no charge, are effectively short circuits. Thus, a surge of displacement current ATTEMPTS to flow. This attempted surge of current is minimized by the self inductive action of the radio frequency choke (RFC). Due to the different values of resistance

A1. No. The series-fed Hartley obtains its feedback by the flow of plate current through the portion of the tank coil labeled L_1. Opening C_2 would permit plate voltage variations to effect oscillator stability.

A2. Through L_1, R_E, emitter-base junction, R_B, RFC, and back to the positive terminal of V_{CC}.

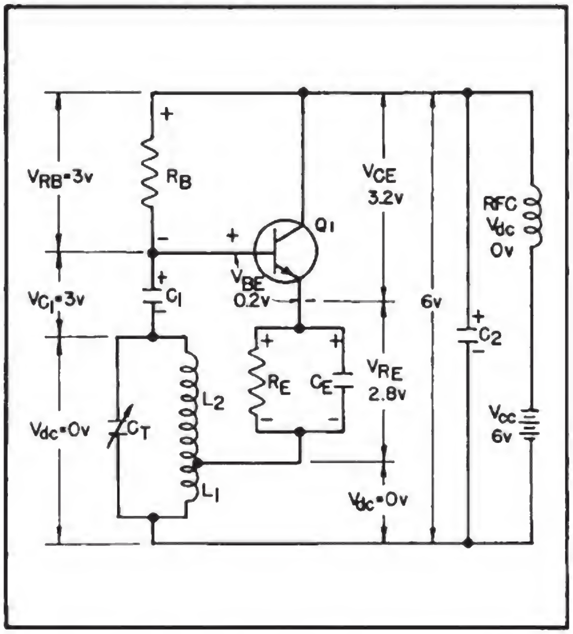

Figure 33-3 - DC voltage distribution.

present in the individual capacitors charge paths each capacitor will have a different value of time constant. However, after a few cycles of operation all of the capacitors will have acquired an average, or dc, value of charge. Also, the resistive components of the circuit will exhibit voltage drops caused by dc biasing currents. The distribution and polarities of these dc voltage drops is shown in Figure 33-3. Representitive values of voltage have been assigned merely to give the reader an approximation of relative magnitudes.

For this explanation the resistance of the coils L_1, L_2, and RFC will be considered to small to give a measurable dc voltage drop, although, in actuallity a very small voltage drop will exist across each one.

Notice that C_2 charges to the value of the supply voltage. While actually very complex, the circuit can be simplified by being thought of as essentially two main branches in parallel with the 6 volts across C_2. The first branch consisting basically of transistor Q_1's collector-to-emitter resistance (r_o) and the stabilizing resistor R_E. The feedback coil L_1 is in series with this branch but since it is considered a dc short (a direct connection with very low resistance) it effectively connects the bottom of the stabilizing resistor to dc ground potential. Notice that the sum of the voltage drops V_{CE} and the stabilizing network V_{R_E} equal 6 volts. The second branch primarily consists of the base biasing resistor R_B and the isolation capacitor C_1. Since coils L_1 and L_2 are considered to be dc shorts they effectively connect one end of C_1 to the dc ground potential. Since capacitor C_1 is in parallel with the stabilizing network and the base-to-emitter junction it will charge to the sum of the voltage drops V_{BE} and V_{R_E}, or 3 volts. An important point to note is that the tank capacitor (C_T) accumulates no measurable average charge because it is shunted by the low resistance of the tank coil. The voltage drop V_{BE} is of such a polarity as to apply forward bias to the transistor.

Q3. Give the main discharge path of capacitor C_E (Figure 33-3) and the purpose of this discharge.

33-5. Positive Alternation of Tank Voltage

Previous chapters have explained the production of an alternating voltage by a tank circuit and the internal transfer of energy between the tank coil and the tank capacitor. The action of the tank circuit in the various transistor oscillators remains the same. Therefore, only the method of replacing dissipated energy in order to sustain oscillations will be mentioned.

As the top of the tank, in Figure 33-3 assumes a positive potential this potential is coupled to the base of Q_1 by capacitor C_1. The positive potential on the base results in an increase of the forward bias of Q_1. The increasing forward bias causes collector current to increase. An increase in collector current means an increase in the emitter current flowing through the feedback winding (L_1) of the tank coil. Increased current flow through L_1 will result in more energy being supplied to the tank circuit, which, in turn, will increase the positive potential at the top of the tank and, hence, the forward bias.

This regenerative action will continue until such time as the combined actions of transistor non-linearity, the effect of R_E, and the dc resistance of L_1 and the power supply cause the rate of change of current through L_1 to decrease. As the rate of change of current through L_1 decreases less energy is supplied to the tank circuit and, through normal tank action, the positive

potential at the top of the tank circuit begins to decrease. This causes a decrease in the forward bias which, in turn, causes the collector and emitter currents to decrease.

At the instant the potential of the tank circuit has decreased to zero the energy of the tank is contained in the magnetic field of the coil, the dc voltage conditions will again be as depicted in Figure 33-3, the tank voltage will have completed its positive alternation, and the oscillator will have completed a half cycle of operation.

Q4. In which component of the tank circuit is the energy "stored" when the tank voltage is at its positive peak?

33-6. Negative Alternation of Tank Voltage

Due to the charging of the tank capacitor in the opposite direction, by the collapsing magnetic field of the coil, the top of the tank circuit will begin to assume a negative potential as the oscillator starts on its second half cycle of operation. This negative potential will be coupled to the base of Q_1 by capacitor C_1 where it will oppose the forward bias of the transistor. Most transistor oscillators are operated class A in order to limit the effect of varying parameters on the stability of the oscillator. Therefore, the value of tank potential coupled to the base, in opposition to the forward bias, will rarely be great enough to cause reverse biasing of the input junction. For instance, in the case of the values established in Figure 33-3 the peak value of the tank signal will not be greater than 0.2 volt. Thus, on the negative half cycle, with a 0.2 volt peak signal and class A operation, the potential coupled to the base from the tank will just cancel the 0.2 volt of forward bias of the transistor. Of coarse, as the negative going tank signal decreases the forward bias, the collector and emitter currents will decrease.

When the tank signal reaches its maximum negative value the collector and emitter currents will be minimum, the magnetic field of the coil will have completely collapsed, and the oscillator will have completed three fourths of a cycle of operation.

At this point the tank capacitor will begin to discharge, thereby, decreasing the negative potential on the top of the tank. As the negative potential coupled to the base decreases the opposition to forward bias will decrease. This will, in effect, cause the forward bias to begin increasing with a resultant increase in emitter current flowing through the feedback coil L_1. As stated previously, an increase in current through L_1 will cause additional energy to be fed to the tank circuit, thereby, replacing energy dissipated within the tank circuit. If the energy replaced is equal to, or greater than, the energy dissipated, oscillations will be sustained.

It was previously mentioned that the operation of the vacuum tube oscillator circuit (Figure 33-1) is generally the same as the operation of the transistor circuit (Figure 33-2). One important difference, however, is the method of obtaining bias. Bias for the vacuum tube circuit depends on the flow of grid current during the positive peaks of the oscillator cycle. The amount of bias depends on the values of the grid leak components used (C_1 and R_g), and the amplitude of the input signal.

In the transistor circuit, bias depends on the amount of current flowing through the base-to-emitter junction. The value of this bias current is determined by the supply voltage (V_{CC}) and the resistance in the bias current path. The path of bias current is from the negative terminal of V_{CC}, through L_1, stabilizing resistor R_E, the base-emitter junction, R_B, RFC, and back to the positive terminal of V_{CC}. The radio frequency coil and L_1 cause no measurable dc voltage drop. Knowing the value of voltage drops across R_B and R_E (caused by the bias current) simplifies the determination of the bias voltage. The bias voltage (V_{BE}) equals the supply voltage (V_{CC}) minus the sum of the other voltage drops. Mathematically:

$$V_{BE} = V_{CC} - (V_{R_B} + V_{R_E})$$

It is usually desirable to operate the transistor circuit class A. It should be recalled from a previous chapter that parameters vary greatly when the transistor is operated on a nonlinear portion of the curve. This would tend towards oscillator instability.

Since the amount of parameter variation is minimized when the point of operation is on the linear portion of the dynamic curve, better oscillator stability can be achieved by class A operation.

Q5. What would be the effect on oscillator operation if the energy replaced in the tank were slightly less than the energy dissipated by the tanks internal resistance?

SHUNT-FED HARTLEY OSCILLATOR

It was previously shown that the main disadvantage of the series-fed Hartley oscillator is the fact that a relatively large value of direct current flows through the feedback portion of the tank coil. This disadvantage can be overcome by a variation of the connection of the feedback coil as shown in Figure 33-4. When connected in this manner the feedback coil is

\3. From the negative plate of the capacitor through R_E to the positive plate of the capacitor. To maintain the voltage across R_E as steady as possible.

\4. The tank capacitor.

\5. Oscillations would gradually die out.

---

solated from the dc component of plate current ›y capacitor C_2. The oscillator derives its ıame from the fact that the feedback component L_1) is connected in shunt rather than series vith the power supply.

The operation of the transistor shunt-fed Hartley is very similar to that of the electron ube shunt-fed Hartley. For this reason a brief ·eview of the electron tube circuit (Figure 33-4) vill be given before the explanation of the transstor version.

33-7. Shunt-Fed Electron Tube Hartley Oscillator

It will be assumed that the oscillator has been n operation for a period of time and that all ca›acitors have assumed their normal operating :harges. The functions of the components are he same as in section 33-2. It will also be ıssumed that at the instant the explanation be;ins the oscillator tube is cut-off by the action ›f grid leak bias, the tank voltage is at the zero ›oint of its cycle, the energy of the tank is ıtored in the magnetic field of the tank coil, and hat the positive alternation is about to begin.

1. It will be noticed that C_2 and L_1 (Figure 33-4) are in parallel with the tube. Since virtually no dc drop occurs across L_1, C_2 will be charged to the value of V_1's plate voltage. Due to the fact that V_1 is cut-off at this time its plate voltage will be equal to E_{bb}.

2. The tank coil will now appear as a generator, by virtue of its field starting to collapse, and cause a clockwise circulating current to flow in the tank. The circulating tank current will begin to charge tank capacitor C_T (with a positive polarity at the top of the tank).

3. As the tank potential becomes increasingly positive a point is reached where this potential will cause the tube to come out of cut-off. At this point tube current begins to flow.

4. The surge of plate current causes plate voltage to be reduced by the action of the radio frequency coil. Since the value of plate voltage is now less than the charge of C_2, C_2 will attempt to force a displacement current around through L_1 and the tube in an effort to equalize its charge.

5. The displacement current through L_1 will cause it to induce a voltage in L_2 of such a polarity as to aid the circulating current which is charging C_T. This is the manner in which dissipated energy is replaced in the tank circuit.

6. When the tank capacitor is charged to its maximum positive potential the coil field will be completely collapsed, and the circuit voltage will be at the peak of its positive alternation.

7. At this point C_T will begin to discharge and the positive potential of the tank will begin to decrease. Tube current will begin to decrease. Plate voltage will increase. When the positive tank potential falls below the value of charge on C_1, grid leak bias action will apply a negative potential to the grid causing the tube to go into a cut-off condition.

8. When the tank potential decreases to zero the oscillator will have completed a half cycle of operation. Flywheel action of the tank (C_T being charged to its maximum value, with a negative potential at the top of the tank, by the coil field and then discharging back to zero) will complete the negative alternation of the operating cycle while the tube remains in a cut-off condition.

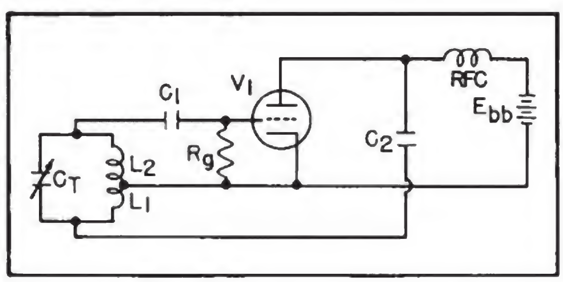

Figure 33-4 - Shunt-fed electron tube Hartley.

Q6. During what part of the cycle of tank voltage does capacitor C_1 (Figure 33-4) receive its charge?

33-8. Shunt-Fed Transistor Hartley Oscillator

Figure 33-5 shows the transistor version of the shunt-fed Hartley oscillator. The functions of the components are the same as described in section 33-3. In addition to providing decoupling between the transistor and the power supply, the radio frequency choke (RFC) also acts as the collector load for Q_1. The RFC developes the variations in collector voltage which are coupled to the feedback coil L_1 by isolation capacitor C_2. The value of C_2 partially determines the amplitude of the feedback signal. If the value of C_2 is relatively large its X_C will be low. A low X_C will cause very little of the feedback voltage to be dropped across the capacitor, thus, providing more feedback to the tank. A relatively small value of C_2 will cause a large amount of the feedback signal to be dropped across its large reactance, thus, reducing the amplitude of the feedback signal to the tank. C_2 also acts as an isolation capacitor to prevent the dc component from flowing through the feedback coil.

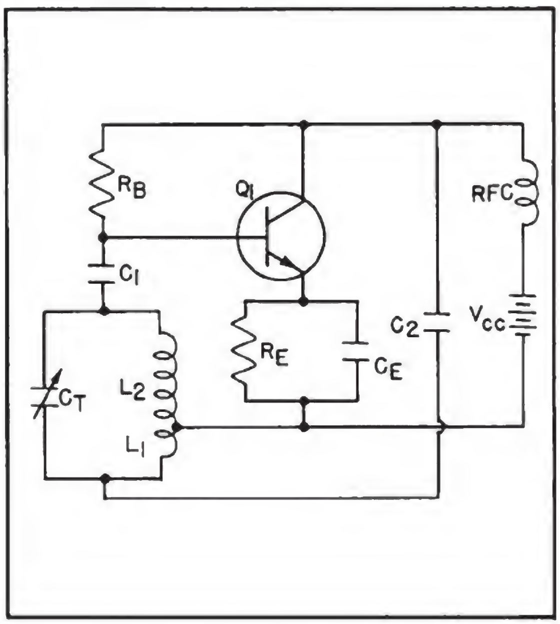

Figure 33-5 - Transistor shunt-fed Hartley.

There is a continuous path for dc bias current through R_E, the base-emitter junction, R_B, and RFC. This allows an operating point to be established for class A operation and also insures the starting of the oscillator.

Aside from these slight differences the action of the tank circuit and feedback mechanism in the transistor shunt-fed Hartley is the same as in its electron tube counterpart.

Q7. If insufficient feedback is developed across L_1, should the value of C_2 (Figure 33-5) be increased or decreased to develop more feedback?

COLPITTS OSCILLATOR

The Colpitts oscillator is another type of LC oscillator (the Hartley and Colpitts are termed LC oscillators because the oscillations take place in a resonant LC tank). Although the basic action of both types of oscillators is the same the method of accomplishing feedback (replacing energy in the tank circuit) is different. Where the Hartley oscillator uses inductive coupling to accomplish feedback the Colpitts oscillator uses capacitive coupling to accomplish this function. In the Colpitts oscillator the tank voltage is divided into two parts by tapping the tank capacitor, or rather by connecting two capacitors in series, instead of the inductor as in the Hartley. The two tank capacitances can be paralleled directly with the input and output interelement capacitances of the tube or transistor, and thereby minimize the effects of these capacitances on the tank circuit operation. Because of this the Colpitts oscillator exhibits slightly better frequency stability than the Hartly.

33-9. Electron Tube Colpitts Oscillator

A detailed discussion of the electron tube Colpitts oscillator was given in section 24-23. Therefore, only a brief explanation will be offered at this time. Figure 33-6 shows the schematic of the Colpitts oscillator to be used in the following discussion.

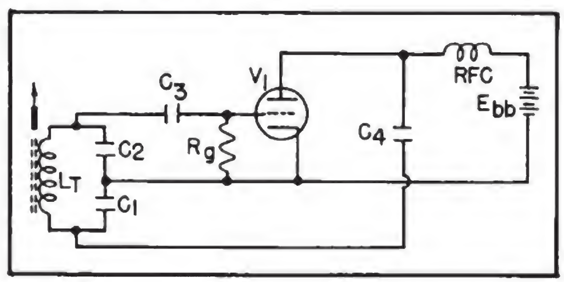

Figure 33-6 - Electron tube Colpitts oscillator.

Notice that the tank circuit is composed of a variable tank inductance (L_T) and tank capacitances C_1 and C_2. The voltage developed across C_2 is the input signal to the tube and the amount of feedback depends on the ratio between C_2 and C_1. For a specific amplitude of signal, the ratio of C_2 to C_1 should remain the same. Since it is easier to use fixed capacitors in this application the Colpitts oscillator lends itself well to the

A6. During the positive alternation when the potential at the top of the tank exceeds the charge of C_1.

A7. Increased. Increasing C_2 will decrease its reactance.

---

use of permeability tuning (indicated by the arrow and dotted core of L_T) in order to adjust the frequency of tank circuit oscillation.

It is assumed for the purpose of explanation that the tube is ready to conduct (filament is hot). Upon application of plate voltage, tube current and a charging current for C_1, C_2, and C_4 will begin to flow. One path for charging current is up through C_2, down through L_T, through C_4, RFC, and back to V_{CC}. This charging current builds a small field around L_T.

1. When the rate of change of charging current begins to decrease, the field of L_T will begin to collapse. This will cause a circulating tank current to begin flowing in a counterclockwise direction. The collapsing field, and the discharge of C_1, increase the charge of C_2 and thereby the positive potential on the top of the tank.

2. The positive tank potential is coupled to the grid by C_3 and causes plate current to increase and plate voltage to decrease.

3. The decreasing plate voltage causes C_4 to move a displacement current around through C_1 and the tube. This displacement current aids the circulating tank current in charging C_1 with a negative potential at the bottom of the tank. This is the manner in which the feedback replaces dissipated tank energy.

4. This action will continue until the field around L_T has completely collapsed. At this point circulating tank current will be zero, plate current will stop increasing, plate voltage will stop decreasing, and the tank voltage will be at the peak of its positive alternation.

5. Capacitors C_1 and C_2 begin to discharge. Circulating tank current reverses direction. The positive tank voltage begins to decrease causing plate current to decrease and plate voltage to increase. C_4 starts to charge aiding tank circuit action.

6. This action will continue until the tube is cut-off by grid leak action. At this time, plate voltage is equal to E_{bb}, C_4 is no longer charging, and the feedback cycle has stopped. Tank oscillation is now maintained by the flywheel effect. In other words, the field of L_T (built up by the discharging tank capacitors) collapses and charges the capacitors with the opposite polarity (negative at the top of the tank). The capacitors will, in turn, discharge and rebuild the field. The above action takes place with the tube in a cut-off condition. When the field is rebuilt the oscillator has completed a full cycle of operation and is now ready to start its second cycle.

33-10. Transistor Colpitts Oscillator

Figure 33-7 illustrates the schematic diagram of a transistor Colpitts oscillator. Like the transistor Hartley, the transistor Colpitts is also operated class A. The value of bias

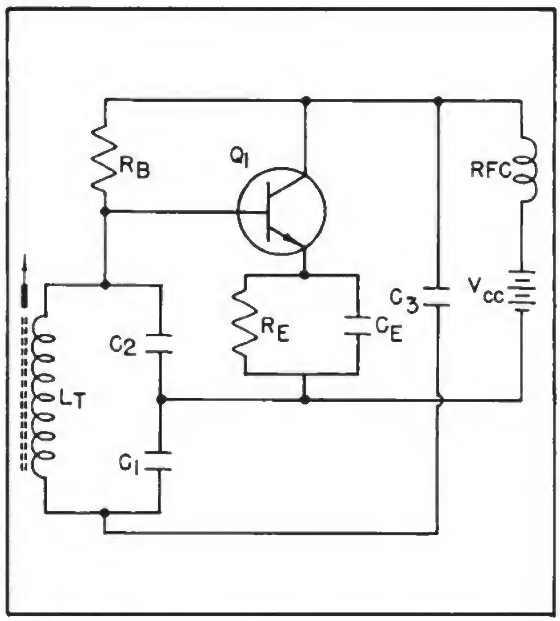

Figure 33-7 - Transistor Colpitts oscillator.

current necessary to establish the desired operating point is determined by the standard voltage divider, formed by R_E, the base-emitter junction, and R_B. It will be noticed that no coupling capacitor is used between the base and tank circuit in Figure 33-7. The main function of this capacitor is to provide dc isolation of the tank coil. In the transistor Colpitts dc isolation of the tank coil is provided by capacitors C_1 and C_3. The basic actions of the tank circuit and

method of obtaining feedback are the same for the transistor version as for its electron tube counterpart.

Q8. What are the functions of capacitor C_3 (Figure 33-7).

SELECTION OF CONFIGURATION

It is possible to use any of the three basic transistor configurations (CC, CB, CE) in the construction of an oscillator. However, certain considerations in the application of the circuit (high operating frequency, large output power required, etc.) usually make the selection of one configuration more desirable than another.

33-11. Common-Collector Configuration

Since there is no phase reversal between the input and the output circuits of a common-collector configuration it is not necessary for the feedback network to provide a phase shift, However, since the voltage gain is less than unity and the power gain is low the common-collector configuration is very seldom used in oscillator circuits.

33-12. Common-Base Configuration

The power gain and voltage gain of the common-base configuration are high enough to give satisfactory operation in an oscillator circuit. The wide range between the input resistance and the output resistance make impedance matching slightly harder to achieve in the common-base circuit than in the common-emitter circuit.

A transistor used in the common-base configuration exhibits better high frequency response than when used in the common-emitter configuration. Therefore, the oscillator will usually be of the common-base type when the operating frequency is equal to, or greater than, the alpha cut-off frequency of the transistor.

It has been stated that a transistor produces no phase shift when used in the common-collector or common-base configurations and 180° phase shift when used in the common-emitter configuration. This is true ONLY when the transistor is operated far below its alpha cut-off frequency. As the frequency of operation approaches the alpha cut-off frequency, the transistor introduces an additional phase shift. At the alpha cut-off frequency this additional phase shift can be as much as 58°. Since common-base oscillators are used when the operating frequency is at, or beyond, the alpha cut-off frequency, this phase shift must be compensated for by the feedback network.

33-13. Common-Emitter Configuration

Due to the high power gain of the common-emitter configuration, and the relatively low frequency of operation (in relation to the 2-10 Mc alpha cut-off of transistors used at broadcast band frequencies), the CE configuration is most often used for broadcast band receiver oscillators.

In order for energy fed back from the output to be in phase with the energy at the input the feedback network of a common-emitter oscillator must provide a phase shift of approximately 180°. Usually a CE transistor oscillator is operated at a frequency sufficiently below the alpha cut-off so that the additional phase shift is negligible. However if an additional phase shift, due to the transistor, is present it must be added to the phase shift provided by the feedback network.

An additional advantage of the common-emitter configuration is the moderate range between the input and the output resistances which simplifies the job of impedance matching.

Q9. List one instance where a CB configuration would be preferred over a CE configuration for an oscillator.

FREQUENCY AND AMPLITUDE STABILITY

The function of a receiver oscillator is to produce a sinusoidal waveshape of a specific frequency and amplitude. In the performance of this function the stability of an oscillator is a very important consideration. Depending on its application an oscillator may be required to have either good FREQUENCY STABILITY or good AMPLITUDE STABILITY, and in many cases both. Of the two, good frequency stability is usually considered more important.

33-14. Frequency Stability

The frequency stability of an oscillator is a measure of the degree to which a constant frequency output is approached. The better the frequency stability, the closer the output will be to a constant frequency.

Frequency instability (variations above and below the ideal constant frequency) may be caused by transistor characteristics or by variations in external circuit elements.

It has been stated that (where output power is not of prime importance) transistor oscillators are often operated class A to insure stability. To this end the dc operating point is chosen so that the operation of the transistor oscillator occurs over the most linear portion of the transistor characteristic curve. When

8. Provide dc isolation and couple collector voltage changes to feedback capacitor C_1.

9. When the operating frequency approaches or surpasses the alpha cut-off frequency of the transistor.

ıe operation of the circuit falls into a non-near portion of the transistor characteristic ırve, because of variations in dc bias volt-ges, the transistor parameters (dc voltages nd currents, etc.) vary. Since these para-ıeters are basic to the transistor and, by ıeir effect on feedback voltages and current, ffect the frequency of tank oscillation, operating ·equency variations will occur with changes ı bias voltage. Thus, a constant supply voltage ı a prime requirement for frequency stability.

The interelement capacitances of the trans-ıtor have a greater effect on frequency stability ıan do the other parameters. The most im-ortant of these interelement capacitances is the ıput capacitance (sometimes called EMITTER IFFUSION capacitance, DIFFUSION capaci-ınce, EMITTER-BASE capacitance, BARRIER ıpacitance, etc.). The important point is that ıis capacitance exists within the transistor be-veen the emitter and the base and, therefore, ıunts any portion of the tank circuit or feedback ɜtwork connected between these elements. The ılue of this capacitance is dependent, among .her things, on temperature, emitter current, ıd frequency. The additional phase shift in a 'ansistor (mentioned previously) is primarily ıused by this diffusion capacitance. Any change ı the operating point, temperature, current, :c., will alter the value of the input capacitance. .nce the input capacitance is in shunt with the edback circuit, or part of the tank, variation ' this capacitance will effect the frequency of scillation. The frequency stability of a tran-.stor oscillator can be imporved by minimizing ıe effect of input capacitance variation. This accomplished by the use of a regulated power ıpply and by the stabilization of emitter current ıd bias voltages (discussed in Chapter 30).

In oscillators operated at high frequencies .ear alpha cut-off) the effect of diffusion ca-ıcitance variations may be minimized by in-ɜrting a relatively large swamping capacitor :ross the emitter-base elements. The total ıpacitance of the two results in a circuit which less sensitive to variations. The added ca-ıcitor may be part of a tuned circuit as in the olpitts oscillator.

The use of a common bias source for both ɔllector and emitter electrodes maintains a ɜlatively constant ratio of the two voltages. In effect, a change in one voltage is somewhat counteracted by the change in the other, since an increased collector voltage causes an in-crease in the oscillating frequency and an in-creased emitter voltage causes a decrease in the oscillating frequency. However, complete compensation is not obtained since the effects on the circuit parameters of each bias voltage differ.

Changes in operating point with changes in temperature are encountered in the transistor oscillator. The effects of changes in temper-ature on the transistor amplifier were given in Chapter 30. Since the transistor oscillator is merely an amplifier with additional circuitry, the means of temperature stabilization are the same for both oscillators and amplifiers.

Q10. What factors influence frequency stability of the transistor oscillator?

33-15. Amplitude Stability

The AMPLITUDE STABILITY of a transistor oscillator indicates the amount by which the output amplitude varies from an ideal constant amplitude.

The same parameters that affected frequency stability will affect amplitude stability. Output amplitude may be maintained relatively constant by insuring that the feedback is large enough, so that the collector current will be limited by the saturation and cut-off regions of the tran-sistor. Operation in this manner makes the output voltage directly proportional to the supply voltage. Thus, regulation of the supply voltage will insure good amplitude stability. Although this type of operation will cause the output wave-form to be slightly distorted, purity of wave-form is not of prime importance in receiver oscillators.

33-16. Tuning

In a superheterodyne receiver, the local oscil-lator frequency will always differ from the de-sired station frequency by an amount equal to the intermediate frequency. The local oscillator is usually tuned above the station frequency, although, in some cases, it may be tuned be-low. When a receiver, having an IF of 456 kc, is tuned to a station at the low end of the broad-cast band (540 kc for example), the local oscil-lator frequency will be:

$$540 \text{ kc} + 456 \text{ kc} = 996 \text{ kc}$$

When the same receiver is tuned to a station at the high end of the band (1600 kc for example),

the local oscillator frequency will be:

1600 kc + 456 kc = 2056 kc

The frequency range of a local oscillator in a superheterodyne receiver, with a 456 kc IF, will be approximately 996 kc to 2056 kc. The tuning component, usually the capacitor, of the tank circuit is mechanically ganged with the tuning component of the input circuit. This allows the local oscillator frequency to be changed as the station frequency is changed, thereby keeping the LO frequency separated, at all times, from the station frequency by an amount equal to the IF. Since the LO frequency follows the selected frequency as tuning occurs, the LO is said to "track" the selected frequency. As was the case with the tuning capacitors in the RF stage, there will be slight electrical differences (due to manufacturing tolerances) in the oscillator tuning capacitors. For this reason, a trimmer capacitor is usually connected in parallel with the oscillator tuning capacitor. Adjustment of the trimmer at the high end of the band insures better tracking.

Most usually, Hartley type oscillators are tuned by means of a variable tank capacitor. This lends itself easily to the ganged tuning mentioned earlier, since the input frequency selector circuit is usually tuned by means of a variable capacitor. The rotors of these variable capacitors can be mechanically connected to a common shaft.

Tracking is more difficult in a Colpitts oscillator when the tank capacitors are made variable. The two capacitors provide impedance matching between the output and the input circuits. The ratio of impedances must be maintained throughout the range of frequencies or impedance mismatch will occur. Resulting in a loss of feedback power. The difficulty encountered in varying Colpitts oscillator frequencies may be reduced by making the tank coil variable, or by making the split capacitors of the tank fixed and inserting a variable tuning capacitor in parallel with the tank. Operation of the Colpitts in this manner will allow the tank frequency to be varied, but will maintain the impedance ratio between the two fixed capacitors.

Q11. What is meant by the term "tracking?"

33-17. Output Coupling

Two common methods of coupling the output of the local oscillator to the mixer stage are transformer and capacitor coupling. The primary requirement of a coupling device is to couple maximum energy to the following stage. To do so, the output impedance of the local oscillator must be matched to the input impedance of the mixer stage.

Examples of capacitive coupling from oscillator tank circuits are shown in Figure 33-8.

The Colpitts oscillator lends itself more readily to capacitive coupling of the output signal than it does to transformer coupling. Although more power is lost in the capacitive coupling circuit, it exhibits a better high frequency response.

To prevent a large signal voltage drop across the coupling capacitor, its reactance must be small compared with the input impedance of the mixer stage. Since the mixer input impedance is usually low, the capacitance value must be high. With the low voltages used, however, the physical size of the capacitor can be kept small.

Examples of transformer coupling are given in Figure 33-9. Notice that the Colpitts oscil-

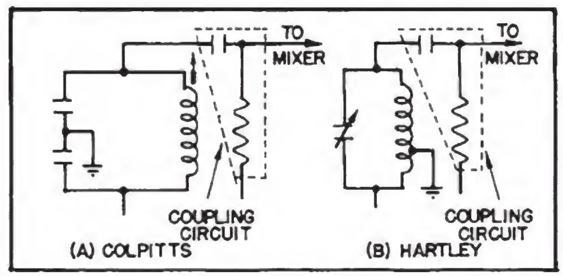

Figure 33-8 - Examples of capacitive coupling.

lator, using transformer coupling, has the two tank capacitors (which establish the impedance ratio) fixed in value, and a tuning capacitor to vary the tank frequency.

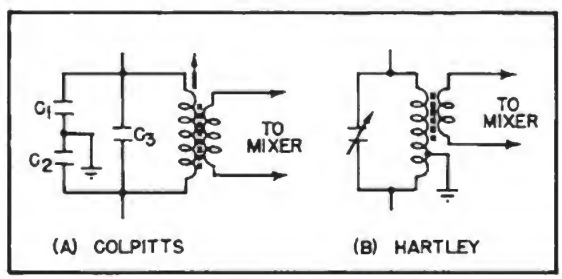

Figure 33-9 - Examples of transformer coupling.

The power efficiency of transformer coupled oscillators is greater than for the capacitive coupled type; however, frequency response is not as good. Transformers are more expensive than the capacitors and resistor used for coupling.

Actually the choice of coupling is determined by many factors such as: operating frequency, output power requirements, available space, type of oscillator, tuning method, etc.

774-639 O - 65 - 7

A10. Emitter current, interelement capacitances, emitter voltage, collector voltage, supply voltage, and the ambient temperature.

A11. The ability of a local oscillator to vary its frequency in step with receiver tuning.

EXERCISE 33

1. What is the purpose of the local oscillator in a superheterodyne receiver?
2. Briefly explain the operation of the basic series-fed transistor Hartley oscillator.
3. Briefly explain the operation of a basic shunt-fed Hartley oscillator.
4. Explain the operation of the basic transistorized Colpitts oscillator.
5. Explain the advantage, in relation to impedance matching, that the CE configuration has for use as an oscillator.
6. What basically determines the frequency of a transistor oscillator?
7. Explain the term "frequency stability."
8. What additional factor must be considered in the determination of a transistor oscillators frequency as the operational frequency is increased?
9. Explain the term "amplitude stability."
10. What is the approximate frequency range of a LO in a broadcast receiver using a 262 kc IF (LO tuned above signal)?
11. Explain the difference between tuning and tracking.
12. What component would normally be adjusted if stations around 600 kc are at the right spot on the dial, but stations around 1500 kc are not on their proper spots?
13. Would it be possible to vary the tank inductance to change frequency in a Hartley oscillator? Explain.
14. Would it be possible to vary both tank inductance and tank capacitance to change frequency in the Hartley oscillator? Explain.
15. If output power is of prime importance, what type of output coupling would most likely be used?

CHAPTER 34

TRANSISTOR MIXERS AND CONVERTERS

The advantages derived from the use of the superheterodyne receiver, as compared to the TRF receiver are uniform gain and selectivity, as the receiver is tuned over a wide range of frequencies. These advantages are possible because in the "superhet" the various incoming radio frequency station signals are reduced to an intermediate frequency having a constant center frequency. This change of frequencies is achieved within the frequency conversion stage of the receiver. There are two basic types of frequency conversion stages used in the "superhet" receiver, one type being the MIXER and the other type being the CONVERTER. The frequency conversion process is the heart of the superheteodyne principle.

This chapter will deal specifically with the mixer and the converter stages used in the transistor receivers; however, much of the theory and many of the characteristics discussed also apply to the corresponding electron tube circuits. For instance, the process of producing new frequencies as the result of applying signals of different frequencies to a nonlinear device is common to both transistor and electron tube circuits. The basic difference between a mixer and converter applies to both the transistor and electron tube circuits; in either case the mixer requires two input signals (radio frequency and oscillator) whereas the converter has its own self contained oscillator and requires only one input signal (radio frequency).

This chapter will cover the block diagram of the mixer, mixer operation and coupling methods, converter characteristics, and converter operation. However, since a thorough knowledge of heterodyning action is necessary in the understanding of mixers and converters, a brief review of the heterodyning process will be presented first.

34-1. Process of Heterodyning

The process of combining two or more frequencies in a nonlinear device and producing new frequencies is called MIXING, HETERODYNING, or FREQUENCY CONVERSION.

The principle of mixing is not related to electronics alone. The basic principles can be traced to Physics.

The production of an audible beat note is a phenomenon which is easily demonstrated. For example, if two adjoining piano keys are struck simultaneously, a note will be produced that rises and falls in intensity at regular intervals. This action results from the fact that the rarefactions and compressions, produced by the vibrating strings, will gradually approach a condition in which they reinforce each other. This occurs at regular intervals of time with an accompaning increase in the intensity of the sound. Likewise, at equal intervals of time, the compressions and rarefactions gradually approach a condition in which they counteract each other and the intensity is periodically reduced. Figure 34-1 illustrates graphically how the resultant difference vibration appears.

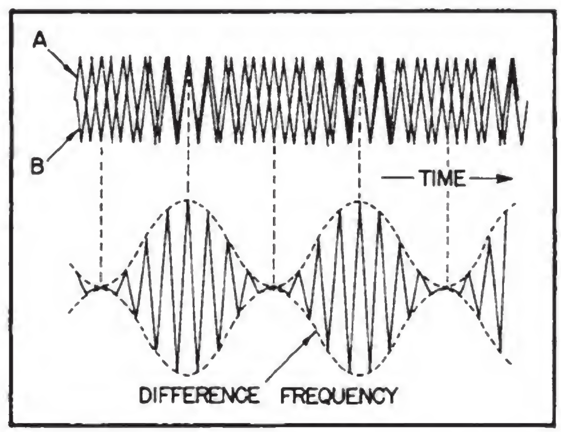

Figure 34-1 - Resultant vibrations produced by two sound waves of slightly different frequency.

If wave (A) were 98 cycles per second and wave (B) were 100 cycles per second, both of the same amplitude, the resultant vibration would rise and fall at a difference frequency of 2 cycles per second. It is important to note that the mere existance of this amplitude variation does not directly indicate the presence of a difference frequency component, in the diagram shown, the difference frequency is observed; but is not necessarily detected until the ear itself acts as a nonlinear device or mixer. To obtain a difference frequency component, it is necessary to

apply the original frequencies to a nonlinear device.

Since the transistor, like the ear, displays non-linear characteristics, the simultaneous application of various frequencies will result in the reproduction of the original frequencies plus the production of various new frequencies.

The necessity for the use of a non-linear device to produce the heterodyning process in a transistor can best be demonstrated by the use of a response curve. The response curve shown in Figure 34-2 is a graph of collector current versus base-emitter voltage. The $V_{BE}—I_C$ curve shown is similar to the $e_c—i_b$ curve of an electron tube. Like the $e_c—i_b$ it is referred to as a dynamic transfer curve. It should be recalled that the transfer curve was constructed, in an earlier chapter, for the purpose of showing the non-linear characteristics of the electron tube amplifier. It was also shown previoulsy (Chapter 26) that simultaneous application of two different frequencies to a linear device produced an output containing only the original frequencies while simultaneous application of the same two frequencies to a non-linear device produced not only the original frequencies but also the sum and the difference of the two original frequencies. Figure 34-2 will be used to explain this action. Waveforms (A) and (B) are used to form the composite waveform (C). Although (C) appears to contain a modulation component at the difference frequency of (A) and (B) the average value of wave (C) at any instant is zero, thus, no useful energy exists at the difference frequency. However, the application of waveform (C) to a non-linear device, such as the transfer curve, causes a heterodyning action between the two original frequencies. This causes the average of the output waveform (D) to be other than zero and to vary at the difference frequency. Although not shown graphically energy is also present at the sum frequency.

If waveform (C) were applied at point Y rather than point X, and the output waveform graphed, it would be noticed that (due to the linear characteristics of the curve in this region (the variations of I_C would be nearly symmetrical. Thus, very little heterodyning action would take place and the average output would be very nearly zero.

It can be concluded that as the device becomes more linear the average amplitude variations in the output approach zero. If the device used were perfectly linear the average output would be exactly zero and only the original frequencies would be present.

Q1. Would heterodyning occur in a device whose output current is directly proportional to its input voltage?

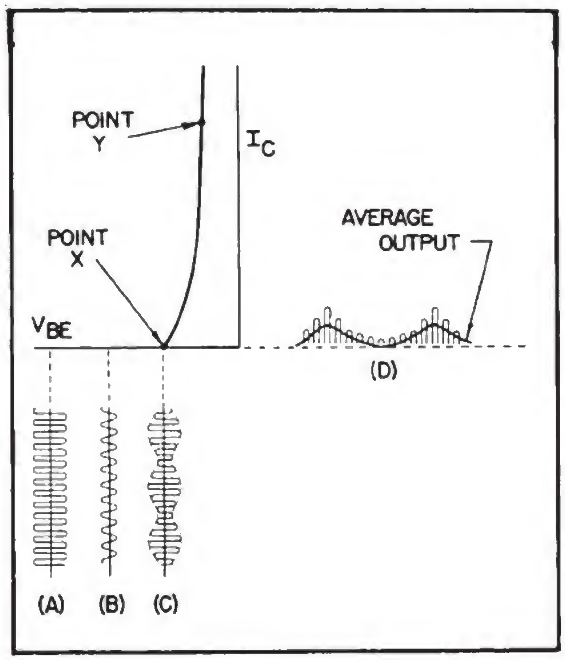

Figure 34-2 - Transfer curve with waveforms

34-2. Block Diagram of a Mixer

The production of beat frequencies in a "superhet" receiver is similar to the examples discussed in Section 34-1. In the mixer stage, the process is entirely electrical and the frequencies are much higher. Figure 34-3 illustrates the block diagram of the mixer stage and the actual frequencies involved (for a specific situation) in the process of mixing.

For simplicity, during the explanation of mixing action, a single 5 kc audio frequency will be used as the original modulating frequency. The frequency of the station carrier signal is 1000 kc. Thus, the AM signal from the transmitter will contain energy at three distinct fre-

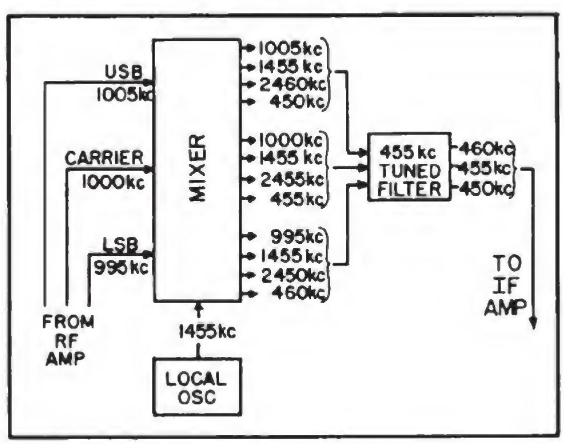

Figure 34-3 - Block diagram of Mixer stage.

quencies, the 1000 kc carrier frequency, the 1005 kc upper sideband frequency, and 995 kc lower sideband frequency. The original 5 kc modulating signal is now contained in the relationship between the carrier and either of the sideband components. This AM signal is used as one of the input signals to the mixer stage and is indicated in Figure 34-3 as the AM signal from the RF amplifier. The other input to the mixer stage is a constant amplitude RF signal from the local oscillator. A prerequisite of the device used as the mixer is that it be non-linear. When using a transistor as a mixer, the curved portion of the dynamic transfer curve is utilized to achieve proper mixing action.

The local oscillator frequency will beat or heterodyne with all the individual components of the modulated wave.

For example, the upper sideband frequency of 1005 kc will heterodyne with the local oscillator frequency of 1455 kc and produce the sum frequency (2460 kc), the difference frequency (450 kc), and the original two (1005 kc and 1455 kc) frequencies. This process will be repeated between the carrier and local oscillator and the lower sideband and local oscillator frequencies. The net result of all these actions is an output which contains twelve significant frequency components. It should be noted that the output will also contain harmonics of all the sum, difference, and original frequencies, however, for simplicity of explanation they will be neglected.

In a practical situation the AM signal would contain many sideband frequencies, which would in turn produce a multitude of significant output frequencies.

Since the output of the mixer contains many frequencies, a tuned resonant tank circuit acting as a filter is employed to select the desired difference frequencies. This filter consists of components so arranged to pass a band of frequencies centered around 455 kc. The response curve of a typical tuned circuit is shown in Figure 34-4. The characteristics of this tank circuit are a relatively uniform gain within the 0.707 points and increasingly less gain for frequencies further from (above or below) the center frequency. Therefore, only frequencies closely associated with the resonant frequency of the tuned circuit will develop sufficient voltage to be passed on to the IF amplifier. Figure 34-4 indicates that the other frequency components present in the mixer output will fall outside the bandpass of the tank circuit and will NOT develop sufficient voltage across the tuned circuit to be passed on to the IF amplifier. Figure 34-4 represents the individual bandwidth of the mixer tank circuit, and not the overall bandwidth of a complete receiver. The half power points have

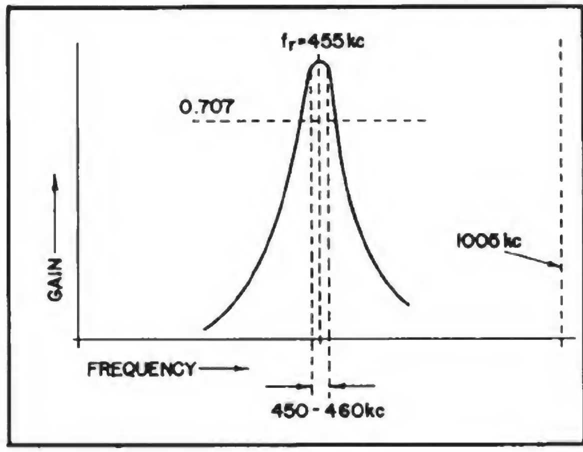

Figure 34-4 - Response curve of Mixer tank

been chosen arbitrarily, and are dependent on receiver circuit design and characteristics.

In the mixer block diagram (Figure 34-3), the output of the tuned circuit contains the frequency components 450 kc, 455 kc, and 460 kc. The output of the tuned circuit constitutes the modulated intermediate frequency signal. It should be noted that, although the sideband and carrier have been converted to lower frequencies, the same relationship between the sideband frequencies and carrier has been maintained. Thus, the original modulating frequency still exists in the IF signal by virtue of the relationship between the carrier and sideband frequencies.

Q2. Why must the frequency response between the half-power points be greater than the highest modulating frequency?

34-3. Transformer Coupled Mixer

It has been stated that the Transistor Mixer has an RF signal input and an input from the local oscillator. These signals may be applied to the mixer in many different ways, depending on circuit configurations, design considerations, etc. Figure 34-5 is a partial schematic of a CE mixer illustrating the injection of the oscillator signal on each of the transistor elements. Transformer T_1 couples the RF signal from the RF amplifier, or in some cases the antenna, to the base of the mixer Q_1. The primary of T_2 and capacitor C_2 form the tuned output circuit of the mixer stage and are resonated at the intermediate frequency. Transformer T_3 couples the oscillator frequency into the mixer and is shown in three possible connections. T_{3A} acts in series with T_1 and couples the oscillator signal to the base of Q_1. T_{3B} couples the oscillator signal to the emitter and controls the conduction of Q_1

A1. No! This device would exhibit a linear characteristic curve and would therefore be unsuitable for mixing.

A2. If it were otherwise, the sidebands representing the higher modulating frequencies would fall outside the 0.707 points and be attenuated.

---

by varying emitter bias. T_{3C} shows the oscillator signal coupled into the collector circuit. Regardless of where the oscillator signal is applied the mixing action described previously remains the same.

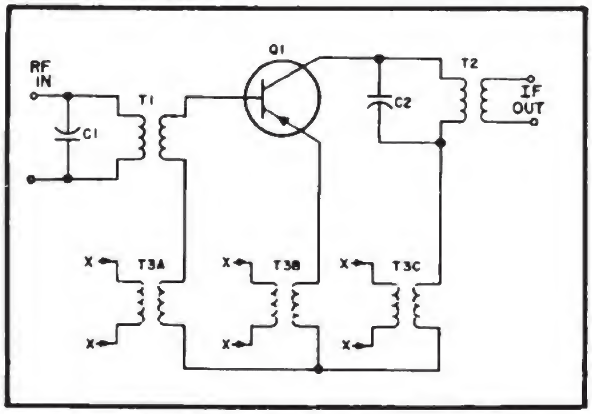

Figure 34-5 - Signal injection methods for Mixer stage.

A complete schematic diagram of a mixer stage employing emitter injection of the oscillator signal is shown in Figure 34-6.

Capacitor C_1 and the primary of T_1 are tuned to the radio frequency signal input. T_1 couples this signal to the base of Q_1. Resistors R_1 and R_3 form a fixed voltage divider network. Their value is chosen to correctly set the magnitude of forward bias for operation on the non-linear portion of the transistors dynamic transfer curve. Resistor R_2 provides bias stabilization and capacitors C_3 and C_4 prevent degeneration. Capacitor C_5 and the primary of T_3 form a parallel resonant circuit for the intermediate (difference) frequency.

34-4. Dynamic Operation

The action of the mixer circuit may be simplified by considering the transistor as being controlled simultaneously by the oscillator signal and the incoming radio frequency signal. The oscillator output is usually much larger (10 or more times) in amplitude than the incoming RF. For this reason, the collector current of the mixer is controlled primarily by the oscillator signal. The signal developed across the emitter-

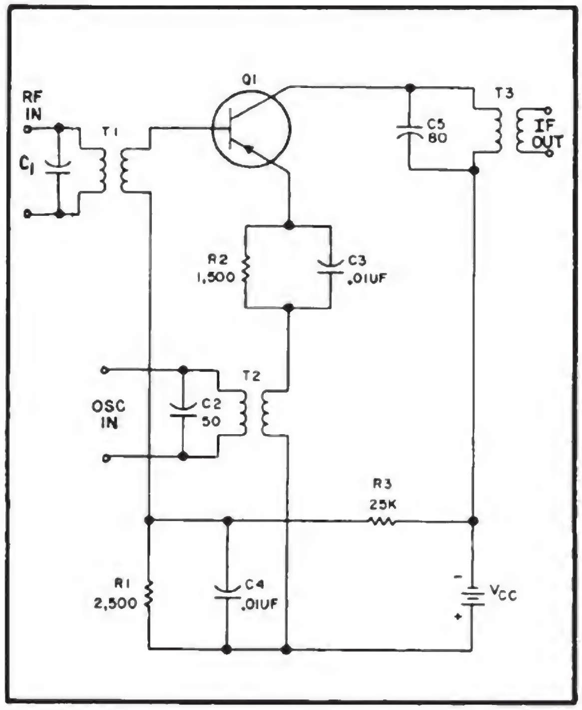

Figure 34-6 - Transformer coupled Mixer.

base junction is the instantaneous sum of the oscillator signal and the radio frequency signal. This emitter-base signal will be a complex waveform due to the heterodyning of the oscillator and radio frequency signals in the non-linear junction of the mixer. The resulting collector current will be a complex waveform consisting of many frequencies including the sum, difference, and the original two frequencies. At resonance, the tank consisting of C_5 and the primary of T_3 will oscillate at, and offer a high impedance to, the difference frequency and its associated sidebands. Because of the high circulating current, the difference frequency will be coupled into the first IF amplifier for further amplification.

Q3. Why is oscillator stability an important prerequisite for mixer operation?

34-5. RC Coupled Mixer

A mixer stage may also employ resistance-capacitance coupling of the local oscillator signal into the stage. An RC coupled mixer stage is illustrated in Figure 34-7.

Capacitor C_3 has a low reactance to RF. C_2 and R_3 couple the modulated RF signal to the base-emitter junction. C_1 and R_2 form an RC coupling network for the local oscillator signal. The primary of T_2 is tuned to select the difference frequency.

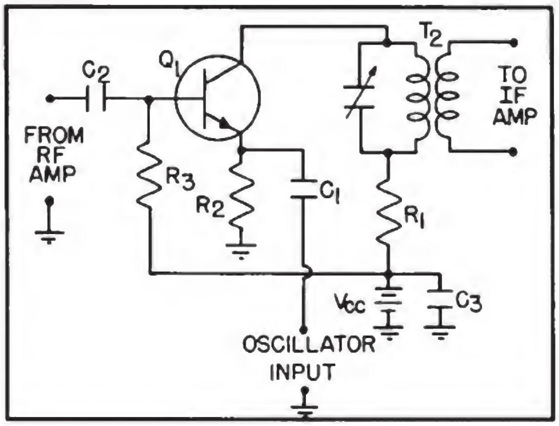

Figure 34-7 - RC coupled Mixer

34-6. Dynamic Operation

The transistor is biased class AB (by R_2, R_3, V_{CC}, and the junction resistance) to utilize its nonlinear characteristics. The local oscillator signal is strong enough to alternately forward and reverse bias the emitter base junction.

Although the conduction time of the transistor is controlled primarily by the oscillator signal the collector current contains components caused by the combination of the RF and oscillator control. Since the mixer stage generates a collector current which contains many frequency components, one must be selected by the output tank circuit. This is accomplished by tuning the output tank circuit to the difference frequency.

The amplification or gain associated with a mixer stage is called the CONVERSION GAIN. This is a comparison of the useful difference frequency energy output to the RF signal energy in the input. Since the useful difference frequency current is just a portion or component of the total collector current signal, the conversion gain is much smaller when comparing it to the gain of the same stage used as an amplifier. Although the mixer or converter stage does not contribute much to the receivers total gain, the function of frequency conversion has been accomplished. Since the output of the mixer stage is always a constant frequency band, the additional gain required can be achieved by the use of fixed tuned IF amplifiers.

Q4. What would be the effect of adding an RF bypass capacitor from emitter to ground in Figure 34-7?

TRANSISTOR CONVERTER

34-7. Characteristics

Sometimes the function of the oscillator and mixer are incorporated into one stage containing two sections. When this is done, the stage is called a CONVERTER. The only external input to the converter stage is the modulated RF signal. The local oscillator signal is generated within the stage. The principle advantage of a converter is the reduction of the number of components and transistors used to accomplish frequency conversion. The disadvantage of a converter is that stable oscillation, which should be a characteristic of the local oscillator, is dependent upon a minimum of nonlinearity, whereas, the frequency conversion process requires a high degree of nonlinearity.

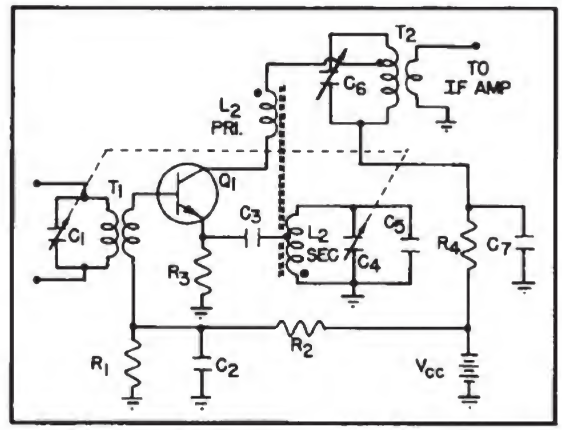

Figure 34-8 - Converter circuit.

A typical converter circuit using an NPN transistor is shown in Figure 34-8. A converter circuit is characterized by only one input signal. The modulated RF signal from the RF amplifier is transformer coupled to the base of Q_1. The local oscillator signal is generated internally by the action of L_2, C_4 and C_5. The circuit may be redrawn to illustrate the oscillator circuit (Figure 34-9), and the Mixer circuit (Figure 34-10) separately.

34-8. Oscillator Section

The oscillator portion of the frequency converter, shown in Figure 34-9, is a basic Armstrong oscillator. Note the positive side of the tickler coil (L_2 primary) has been connected directly to V_{CC} to show oscillator action. The components T_2, C_6, R_4, and C_7, shown in Figure 34-8, will be discussed in connection with the operation of the mixer section of the converter. R_1 and R_2 form a voltage divider network to establish the transistor operating point. R_3 provides emitter stabilization, and is essentially bypassed to ground for RF signals through a few turns of the secondary winding L_2 and the capacitor C_3. R_3, although bypassed for RF, must develope a voltage at the oscillator fre-

A3. To insure that the resultant IF produced and applied to the IF stage will be a constant frequency.

A4. Bypassing R_2 would prevent the development of the local oscillator control signal.

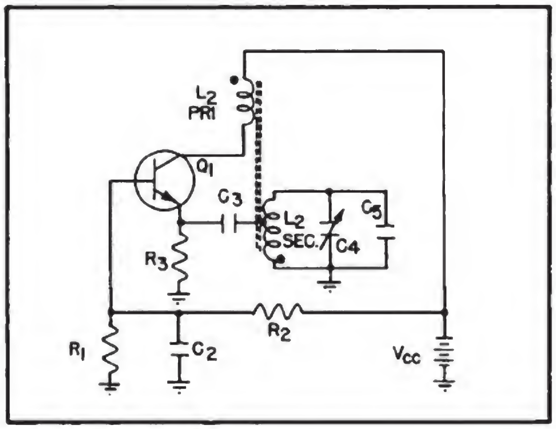

Figure 34-9 - Oscillator section of converter.

quency in order to control the emitter bias. Careful observation will show that the local oscillator section of the converter is a CB configuration, due to the fact that the base is common to both the emitter and collector circuits. C_2 is a bypass capacitor around R_1 to prevent degeneration and effectively places the base at ac ground.

34-9. Operation of the Oscillator Circuit

Oscillations in this circuit are started the moment the dc power (V_{CC}) is applied to the circuit. At that moment, a surge of current flows through the transistor, and the tank circuit goes into oscillation. The amplified oscillations appear across the L_2 primary, and are inductively coupled to the L_2 secondary. The feedback signal is regenerative and of sufficient magnitude to sustain oscillations. The secondary of L_2 is tapped to achieve a proper impedance match between the high impedance tank circuit and the relatively low impedance of the emitter circuit.

34-10. Mixer Section

Figure 34-10 illustrates the mixer section of the converter. For simplicity, the components of the oscillator section have been replaced by the block labeled "LO tank". The modulated RF signal is transformer coupled to the base of the converter. The constant amplitude local oscillator signal is coupled through C_3 to the emitter of the transistor. Transistor Q_1 is biased by R_1 and R_2 on the nonlinear portion of the dynamic characteristic curve for proper heterodyning action. The transistors collector current is controlled simultaneously by the oscillator signal and the incoming signal. The collector current will be a complex waveform consisting of many frequencies including: the local oscillator signal frequency, the received incoming signal frequency, the sum of these two frequencies, and the difference between these two frequencies.

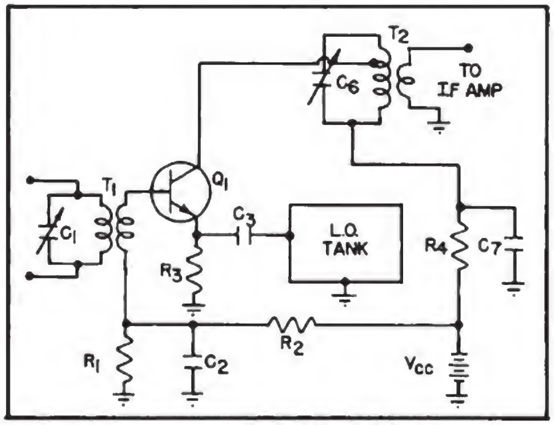

Figure 34-10 - Mixer section of converter.

The primary of the output transformer and C_6 form a resonant circuit, which is tuned to the difference frequency (IF). Thus, the IF frequency will develop a relatively large voltage across the tuned circuit while the remaining signals (LO, RF and the sum frequency) will develop relatively small voltages across the tuned circuit. R_4 and C_7 serve as a decoupling network to filter out the remaining energy of the undesirable signals.

Q5. Why are tapped coils used on transistor circuits?

Q6. What would be the result if the tickler coil were to become shorted?

34-11. Image Frequency

IMAGE FREQUENCY is defined as an interfering transmitted signal the frequency of which always differs from the desired station frequency by twice the IF. In other words, image frequency = station frequency $\pm$ (2 X intermediate frequency). The plus sign of the formula is used if the local oscillator frequency tracks (operates) above the station frequency. The minus sign of the formula is used if the local oscillator tracks below the station frequency. The oscillator tracking below the station frequency is generally used for the higher frequency bands, and the

oscillator tracking above the station frequency is generally used for the lower frequency bands such as the broadcast band.

The frequency relationships between image frequency, oscillator frequency, and station frequency, versus receiver frequency are illustrated in Figure 34-11.

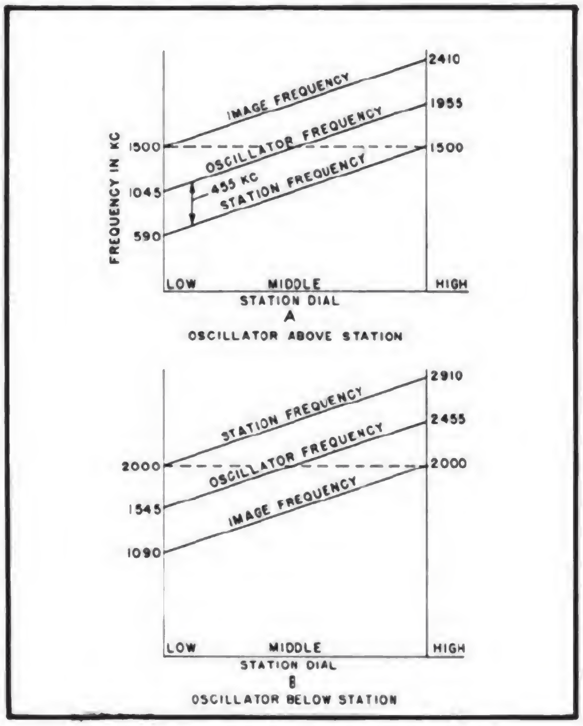

Figure 34-11 - Relationship of image frequency to station frequency

As an example, if a receiver having an intermediate frequency of 455 kc is tuned to receive a station frequency of 1500 kc, its local oscillator will be at a frequency of 1955 kc. Provided a station is operating at 2410 kc and is of sufficient strength, it too will beat with the 1955 kc oscillator signal and produce a difference frequency of 455 kc. This image station is out of the broadcast range but could consist of other types of transmissions such as police calls, marine signals, etc. Both stations when mixed with the local oscillator will produce the correct difference frequencies. Thus the IF amplifier cannot separate the desired station from the image signal. Since the 1500 kc station and the 2410 kc image are heard simultaneously in the output of the receiver, distortion and interference result.

The selectivity of the preselector (RF amplifier) and antenna circuits tends to reduce the strength of these images or unwanted signals prior to entering the mixer or converter stage.

The ratio of the amplitude of the desired station signal to that of the image is called the IMAGE REJECTION RATIO. In other terms:

$$\text{IMR} = \frac{\text{amplitude of desired signal}}{\text{amplitude of image signal}}$$

For example, if the amplitude of the desired station frequency is equal to 10 microvolts, and the amplitude of the image frequency is 10 microvolts, then the ratio is 1:1. This is undesirable because the image signal being received is as strong as the desired signal. Ratios of 5:1 or 10:1 are very desirable because the amplitude of the desired signal is 5 to 10 times the amplitude of the image signal.

Q7. How can image frequencies be reduced in superhet receivers?

Q8. Why are image frequencies undesirable?

Q9. With a station frequency of 810 kc and an image of 1322 kc, what is the LO frequency?

Q10. Why is an image rejection ratio of 1:5 undesirable?

34-12. Tracking

The oscillator frequency must always differ from the station frequency by an amount equal to the intermediate frequency. When tuning the receiver, the oscillators frequency must change by the same amount as the incoming RF if the difference between them is to remain the same. When this is done the oscillator is said to "track" the RF. Varying both the main tuning capacitor and the oscillator capacitor is accomplished by use of a ganged capacitor.

For proper operation of the receiver the IF frequency must remain constant over the entire tuning range of the receiver. In other words, if the IF of a particular receiver is 455 kc, then, the IF must be 455 kc when the receiver is tuned to the low end of the range (540 kc) as well as the high end of the range (1600 kc).

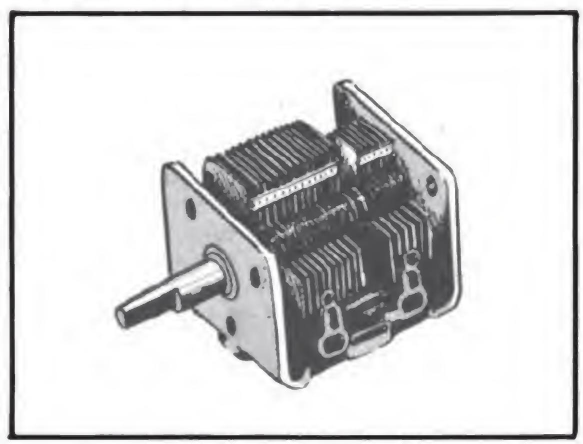

Figure 34-12 - Ganged capacitors.

A5. To achieve proper impedance matching for maximum transfer of energy.

A6. No output from the converter stage because the LO would cease to operate due to insufficient feedback. The only signal present in the collector tank circuit would be the modulated RF frequencies which fall outside the bandpass and receive very little amplification.

A7. The addition of RF amplifiers and tank circuits selective enough to discriminate against images.

A8. Interference and distortion in the output would result.

A9. 1066 kc.

A10. This ratio indicates that the amplitude of the image is 5 times greater than the amplitude of the desired station signal.

Due to manufacturing tolerances, design problems, etc., there is usually some variation in the characteristics of individual capacitors, resulting in improper tracking. A direct result of improper tracking is IF variation. If the RF and oscillator frequencies do not change equally the IF will vary over the tuning range. This means the IF might be 450 kc at the low end of the range, 455 kc at the middle, and 460 kc at the upper limit. Since the IF amplifiers have a fixed bandwidth with a specific center frequency this would result in reduced receiver gain at the high and low ends of the range with possible distortion.

Several methods have been developed to reduce IF variations due to improper tracking. One of these methods is the use of a TRIMMER CAPACITOR. The trimmer capacitor is usually of the compression mica type, and is mounted on the top of the main tuning capacitor as shown in Figure 34-12. Electrically, the trimmers are in parallel with each section (oscillator and RF). The schematic diagram of the oscillator capacitor and trimmer is shown in Figure 34-13. The value of the trimmer capacitor is usually smaller than its associated capacitor.

The trimmer is used to compensate for tracking variations at the high end of the tuning range. The reason for this can be demonstrated by the following example.

NOTE: Values used in the examples are taken from a typical ganged tuning capacitor used in broadcast band receivers. An oscillator capacitor, having a range of 10.6 pf to 172.6 pf, is paralleled with a trimmer, having a range of 2 pf to 17 pf.

To determine the effect of trimmer variation

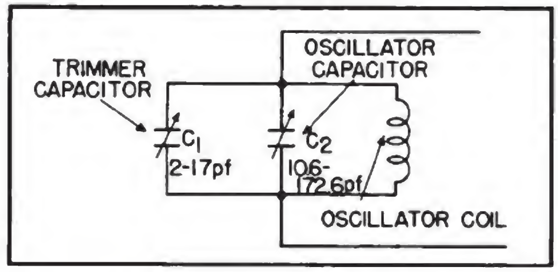

Figure 34-13 - Oscillator tuning capacitor with trimmer.

on total tank capacitance at the high frequency end of the tuning range, set both capacitors at minimum (C_2 = 10.6 pf and C_1 = 2 pf, Figure 34-13). Thus, the minimum capacitance of the parallel combination will be:

$$2 + 10.6 = 12.6 \text{ pf}$$

Maintaining the oscillator capacitor at minimum and setting the trimmer to maximum (C_1 = 17 pf) will cause the minimum capacitance of the combination to increase to:

$$17 + 10.6 = 27.6 \text{ pf}$$

Thus, at the high frequency end of the range the trimmer capacitor can cause a change in total tank capacitance of:

$$27.6 - 12.6 = 15 \text{ pf}$$

Dividing the change in capacitance by the original minimum capacitance and multiplying by a 100 will result in the percentage of change in tank capacitance caused by the trimmer.

$$\frac{15}{12.6} \times 100 = 119\%$$

The effect of trimmer variation on total tank capacitance at the low end of the tuning range is determined by setting the oscillator capacitor to maximum and the trimmer to minimum (C_1 = 2 pf and C_2 = 172.6 pf). The total tank capacitance with these settings will be:

$$2 + 172.6 = 174.6 \text{ pf}$$

Holding the oscillator capacitor at maximum and setting the trimmer at maximum will increase the tank capacitance to:

$$17 + 172.6 = 189.6 \text{ pf}$$

The change in tank capacitance at the low frequency end caused by varying the trimmer will be:

$$189.6 - 174.6 = 15 \text{ pf}$$

The change in capacitance of 15 pf is the same as occurred at the high end, however, the percentage of change is vastly different:

$$\frac{15}{174.6} \times 100 = 8.5\%$$

Therefore, it can be seen that varying the trimmer capacitor effects the percentage of tank capacitance to a much greater extent (119%) at the high end of the tuning range and has relatively little effect (8.5%) at the low end. For this reason, the trimmer is adjusted at the high frequency end of the receivers tuning range.

A method of reducing improper tracking at the low end of the tuning range is by the use of a PADDER capacitor. The padder is usually found only in the oscillator tank circuit. If the oscillator coil has an adjustable powered iron core, the padder is usually not incorporated. The padder is electrically in seiies with the tank circuit. A schematic showing the placement of the padder is shown in Figure 34-14.

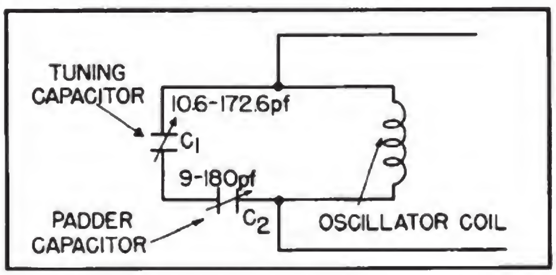

Figure 34-14 - Oscillator tuning capacitor with padder.

The padder is usually a compression type mica capacitor. It has more plate area than the trimmer, and therefore, has more capacity. The capacity value of the padder is approximately the same as the main tuning capacitor. When capacitors are in series the formula:

$CT = \dfrac{C_1\ C_2}{C_1 + C_2}$ must be used to find total circuit capacity.

With the oscillator tuning capacitor set for a minimum capacitance (high frequency) of 10.6 pf, the padder may be adjusted from 9-180 pf. Using the formula, the total tank capacity change will be from 4.8 pf to 10 pf. With the tuning capacitor set for maximum capacity (low frequency) of 172.6 pf, the padder may be adjusted from 9-180 pf. The total circuit capacity change will be from approximately 8.5 pf to 88 pf. Therefore, a change of only 5.2 pf (10 - 4.8) occurs at the high frequency end, whereas, a change of 79.5 pf (88 - 8.5) occurs at the low frequency end of the band. Since the change at low frequency end is approximately 15 times greater than that at the high frequency end, the padder adjustment has the greatest effect at the low end of the tuning range.

An alternate method used to improve the tracking at the low frequency end of the band is to cut slots in the end rotor plates of the oscillator tuning capacitor as shown in Figure 34-15. The capacitance may be varied by bending the individual plates on the rotor. If the plate is bent toward the stator, the capacitance is increased due to decreasing the distance between the plates. If it is bent away from the stator, the capacitance will decrease.

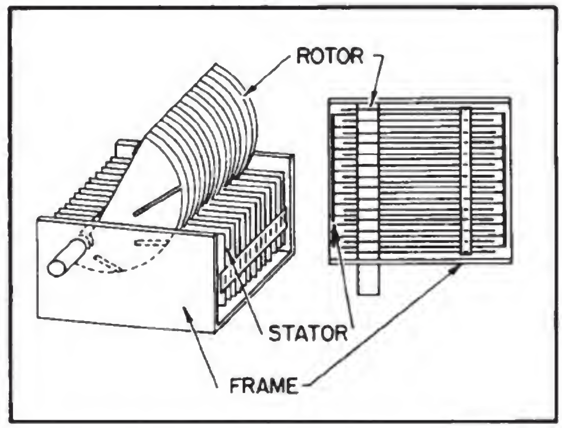

Figure 34-15 - A slotted rotor plate.

The exact frequencies at which the padder and trimmer capacitors are adjusted to improve tracking is usually given in the alignment specifications for the specific receiver.

Q11. When the tuning capacitor is set to minimum capacitance will the plates be fully meshed or fully open? Explain.

Q12. Are trimmer and padder capacitors necessary in the output circuit of the mixer or converter stage? Explain.

A11. Fully open. One of the factors determining capacitance is the plate area. With the plates open plate area is minimum.

A12. No. This tank is tuned to a fixed intermediate frequency.

EXERCISE 34

1. Why is it desirable to convert the various incoming RF to a fixed lower frequency?
2. Why is a knowledge of heterodyning action necessary in the study of the mixer and converter stages?
3. Since the response of the human ear very seldom goes below 30 cps explain how it would be possible for the ear to detect a 5 cps fluctuation in sound intensity.
4. Explain why it is impossible to obtain energy at the difference frequency when two signals are mixed across a perfectly linear device
5. Are audio frequencies present as such in the signal transmitted from the station? Explain?
6. List the sideband frequencies necessary for an RF signal to contain intellegence at 15 kc. Assume a carrier frequency of 650 kc.
7. Assume the RF signal of question 6 is applied to a mixer and reduced to a 462 kc IF signal. What sideband frequencies must be present in the IF signal to preserve the 15 kc intellegence?
8. How does conversion gain differ from the regular gain of a transistor? Explain.
9. If oscillator stability is an important requirement in a certain application would frequency conversion be accomplished by a mixer or a converter stage? Explain.
10. What effect will image frequencies have on a TRF receiver?
11. Define tracking and explain the effect of poor tracking on receiver operation.
12. What would the IMR of a receiver be if the desired signal were 30 microvolts and the image signal were 180 microvolts?
13. What would be the percentage of tank capacity change, due to trimmer variation, at the high frequency end of the range be if a tuning capacitor with a 30-300 pf range and a trimmer with a 3-30 pf range were used?
14. Explain how a slotted rotor plate can be used in place of a padder.
15. Would it be possible for an oscillator tank to contain both a trimmer and padder capacitor?

CHAPTER 35

TRANSISTOR IF AMPLIFIERS

The receiver circuits studied, thus far, have included the RF amplifier, the oscillator, the mixer, and the converter circuits. The amplitude of the signal from the mixer (or converter) is still comparatively weak. Due to its small amplitude, it is not considered practical to feed this signal directly to a detector stage for demodulation. For this reason the superheterodyne receiver includes one or more stages of INTERMEDIATE FREQUENCY amplification between the mixer and the detector stages.

In the block diagram of the superheterodyne receiver (Chapter 31) it can be seen that the IF amplifier block receives its input signal from the mixer block. This signal is an amplitude modulated wave at a frequency somewhat lower than the received station signal. The IF amplifier stage, or stages, produces an amplified reproduction of the mixer output signal (a band of frequencies centered around the IF frequency) and applies it to the detector stage.

In many ways the operation of the IF amplifier is similar to that of the RF amplifier. The signals being amplified, however, are at a lower frequency than for the RF amplifier. Unlike the RF tuned circuits (whose frequency is variable over a wide range), the tuned circuits used in IF amplifiers are fixed at a definite resonant frequency; adjustable tank components being incorporated for alignment purposes only. Since they operate at a fixed band of frequencies, the IF amplifiers can be designed to provide optimum gain and bandwidth characteristics.

This chapter will consider the operational theory of transistor IF amplifiers, including factors affecting the choice of an intermediate frequency and the choice of circuit configuration. Transformers used for interstage coupling at the IF frequencies will be given a detailed explanation. Finally, the principles of single tuning, double tuning, and stagger tuning will be presented.

35-1. Choice of IF Frequency

Many factors are involved in the choice of a receivers IF frequency. In most cases the chosen value is a compromise and may vary from one brand of receiver to another. The value of IF frequency for a specific receiver is chosen by the designer, or manufacturer, from a wide range of values. In fact, since the discovery of the heterodyne principle, IF frequencies ranging in value from 130 kc to 485 kc have been used in broadcast band "superhet" receivers. Values at the low end of this range (175 kc, 262 kc, etc) are used occasionally. The use of a low value IF frequency results in slightly better gain and stability characteristics. A low IF frequency also results in a 2% to 3% improvement in selectivity. However, the advantages gained are overshadowed by the increased susceptibility to image-frequency reception. This occurs because lowering the value of IF frequency moves the image-frequency closer to the frequency of the desired station, thus increasing the possibility of an undesired station having the proper value of carrier frequency to beat with the local oscillator of the receiver and produce a signal at the receivers IF frequency.

The majority of modern broadcast band receivers employ one of three IF frequency values: 455 kc, 456 kc, or 465 kc. These values have been found, through experience, to give sufficient gain and stability characteristics while retaining acceptable image-frequency rejection.

Q1. Would the image-frequency of a receiver tuned to 800 kc, and having an IF frequency of 175 kc, lie within the broadcast band? The local oscillator is tuned above the station frequency.

Q2. Would the image-frequency of a receiver tuned to 800 kc and having an IF frequency of 465 kc lie within the broadcast band?

35-2. Single Stage IF Frequency Amplifier

Before entering into a discussion of the components used in the IF amplifier, a brief, overall description of a typical IF amplifier stage will be presented. Figure 35-1 illustrates a typical transistor IF amplifier. The tuned input circuit, which is formed by the primary of T_1 and capacitor C_1, is actually the output tank circuit of the mixer stage. As previously stated this circuit is tuned to the receiver IF frequency (456 kc in this case). The primary of T_1 is

A1. Yes. The image-frequency would be 1150 kc.

A2. No. The image-frequency would be 1730 kc.

---

tuned by the adjustment of a powdered iron core (indicated by the arrow and slug above the primary coil), while the secondary is untuned. Tapping the primary winding permits matching the impedance of the tuned circuit to the output impedance of the mixer stage. The turns ratio and the coefficient of coupling is chosen so that the impedance of the secondary winding will match the relatively low input impedance of the transistor Q_1.

The operating point for class A operation of the CE configuration is established by a standard voltage divider and bias stabilizing network, consisting of R_1 and R_3. Bypass capacitors C_3, C_4, and C_6 perform their usual function. Emitter bias and emitter current stabilization is provided by R_2. Regeneration is prevented by neutralizing capacitor C_2.

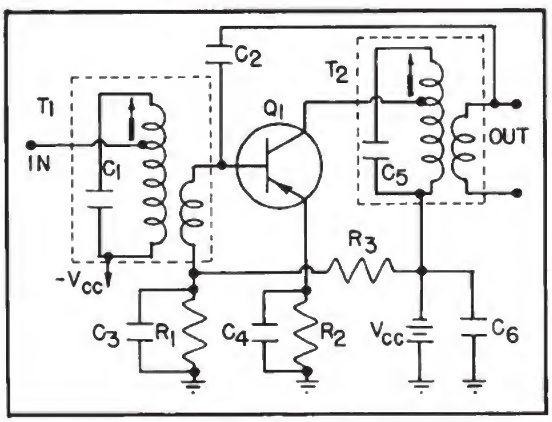

Figure 35-1 - Single stage IF amplifier.

The transistor Q_1 is a small signal, germanium PNP type. In accordance with standard practice, Q_1 is operated in the CE configuration in order to take advantage of the higher power gain.

Transformer T_2 has its primary tuned to 456 kc and an untuned secondary. The primary winding is tapped to prevent the moderate output impedance of Q_1 from shunting the tuned circuit and lowering its Q (explained in the next section). The secondary of T_2 may feed either the detector stage or another IF amplifier stage.

In normal operation, a single stage of IF frequency amplification is impractical due to poor selectivity and poor detector efficiency at low signal levels. However, where cost is a factor and peak performance is not of prime importance it is possible that a single stage of IF frequency amplification may be used.

In most cases the gain, bandwidth, and selectivity requirements of a broadcast band receiver can be satisfied by cascading (connecting in series) two IF stages. When additional gain is needed, it is usually acquired through the addition of audio amplifiers rather than IF amplifiers because of the instability resulting when tuned stages are cascaded.

Q3. If the signal induced in the secondary of T_1 (Figure 35-1) increased in a positive direction, would the voltage drop across R_2 increase, decrease, or remain relatively constant?

Q4. What effect would an increase in ambient temperature have on the voltage drop across the bias stabilizing resistor R_3 (Figure 35-1)?

INTERMEDIATE FREQUENCY TRANSFORMERS

The characteristics and physical construction of transformers used as coupling devices at high frequencies are quite different than for transformers used at audio frequencies. The purpose of this section is to discuss the characteristics and use of interstage transformers, at intermediate frequencies, in transistor receivers.

35-3. Characteristics of IF Transformers

The primary, or the secondary, (and in some cases both) circuits of IF transformers are tuned to a specific resonant frequency.

For any specific resonant frequency the LC product of a tuned circuit is a constant. Thus, the product of the inductance (L) and the capacitance (C), which form the tuned primary of a 455 kc IF transformer, will be identical whether the transformer is used in an electron tube receiver or a transistor receiver. However, due to the magnitude of currents and voltages used in transistor receivers, a transistor IF transformer is considerably smaller in physical size than its electron tube counterpart. Figure 35-2 illustrates the relative size of the IF transformers used in the two types of receivers.

Since the selectivity of a receiver is primarily determined by the IF stages, it is desirable that the tuned circuits (IF transformers) have a relatively high Q. It was shown, in Chapter 9, that the core material is one of the factors having a pronounced effect on the induct-

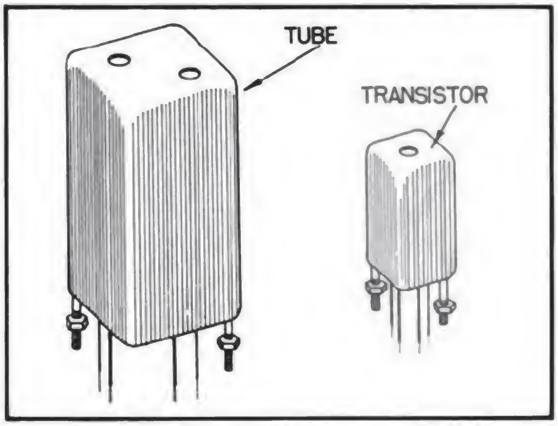

Figure 35-2 - IF transformers.

ance of a coil. Increasing the permeability of the core material increases the inductance of the coil. Use of a powdered iron core (indicated by the dotted lines between the windings as shown in Figure 35-3), in a miniature IF transformer, increases the permeability and allows the required amount of inductance to be obtained with relatively few turns of wire.

Miniature IF transformers are constructed so that the powdered iron core may be moved in and out of the area enclosed by the windings. This effectively varies the permeability of the core, thereby varying the inductance. For this reason miniature IF transformers are referred to as being PERMEABILITY TUNED. When a transformer is permeability tuned the capacitance of the tank is usually a fixed value, however, it is possible to have BOTH capacitive and permeability tuning.

The Q of the transformer with no external load connected is called the UNLOADED Q. When used as an interstage coupling device the primary of the transformer will be shunted by the output resistance (R_o) of the preceding stage and the secondary will be shunted by the input resistance (R_i) of the following stage. This situation is illustrated in Figure 35-3. The Q of the transformer with R_o and R_i connected is called the LOADED Q.

One of the functions of an interstage transformer is to match the relatively low resistance R_i to the relatively large resistance R_o. This can be accomplished by selecting the proper turns ratio of the transformer primary (L_p) to the transformer secondary (L_s). Stated in the form of an equation:

$\frac{N_1}{N_2}$ = the transformer turns ratio

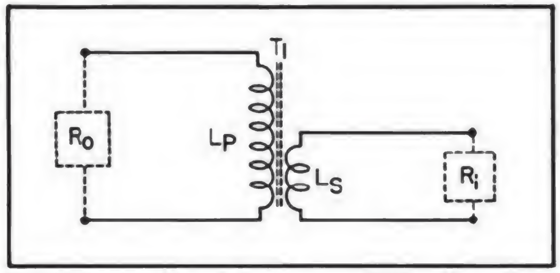

Figure 35-3 - IF transformer - impedance matching.

where: N_1 = the number of turns in the primary winding (L_p)
N_2 = the number of turns in the secondary winding (L_s)

The turns ratio of a transformer may also be stated in terms of the output and input resistances of the transistors.

$$\frac{N_1}{N_2} = \sqrt{R_o/R_i} \qquad (35\text{-}1)$$

where: R_o = the output resistance (connected across primary) in ohms.
R_i = the input resistance (connected across secondary) in ohms

NOTE: The use of this equation assumes unity coupling, which is permissible with a powdered iron core transformer.

In order to demonstrate the use of an IF transformer as an interstage coupling device (and illustrate the effect of shunt resistances, reflected resistances, losses, circuit requirements, etc., on its configuration and operation) an example problem will be examined.

Example problem:

It is desired to use transformer coupling between two, CE, transistor IF amplifiers. The input resistance (R_i) of both transistors is 2 k ohms. The output resistance (R_o) of both transistors is 40 k ohms. The frequency of operation of the IF stage is 450 kc. It was shown, in section 32-5, that when tuned circuits are cascaded the overall bandwidth becomes considerably less than that of any single tuned circuit. Since the overall bandwidth of the receiver should be approximately 10 kc the value of bandwidth for the tuned circuit in the example problem is arbitrarily chosen as 20 kc.

A3. Remain relatively constant due to the action of C_4.

A4. Increase the E_{R3} drop due to more current being drawn by the base-to-emitter junction of Q_1.

Determine the transformer turns ratio:

$$\frac{N_1}{N_2} = \sqrt{R_o/R_i} \qquad (35\text{-}1)$$

$$\frac{N_1}{N_2} = \sqrt{\frac{40 \times 10^3}{2 \times 10^3}}$$

$$\frac{N_1}{N_2} = \sqrt{20}$$

$$\frac{N_1}{N_2} = 4.47$$

Thus, in order to match the impedance of the two CE configurations, and thereby achieve a maximum transfer of power, a transformer having a turns ratio of 4.47:1 should be used.

Figure 35-4 shows a partial schematic (bias supplies, etc., are omitted) of the two CE transistors and the interstage transformer. The circuit symbols used in Figure 35-4 are:

R_i = the input resistance of Q_2.

R'_i = the input resistance of Q_2 reflected into the primary.

R_o = the output resistance of Q_1.

L_p = the primary inductance of T_1.

L_s = the secondary inductance of T_1.

Due to the impedance matching properties of the transformer, the R_i of 2 k ohms is made to appear in the primary as a parallel resistance R'_i of 40 k ohms. Since $R'_i = R_o$, impedance matching has been accomplished. This will effectively lower the output resistance to half of its original value, or approximately 20 k ohms (product over the sum method).

Since the transformer coupling of the example problem is being used in an IF stage, it must contain a resonant circuit. In order to determine the value of capacitance needed to resonate with the primary inductance, at the IF frequency

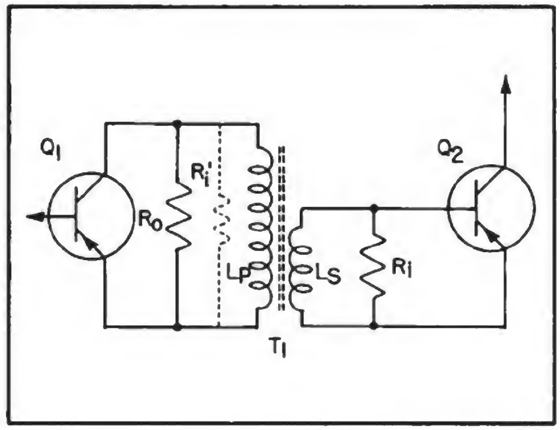

Figure 35-4 - Partial schematic of interstage coupling network.

of 450 kc, L_p must first be determined. The reactance of the primary (X_{Lp}) can be determined by the use of the following equation (the derivation of which is included at the end of this chapter).

$$X_{Lp} = \frac{R_o\,(Q_c - Q_T)}{2\,Q_c\,Q_T} \qquad (35\text{-}2)$$

where: X_{Lp} = reactance of the transformer primary in ohms.

R_o = output resistance in ohms (use of this equation assumes $R'_i = R_o$).

Q_c = UNLOADED Q of the transformer primary.

Q_T = total Q of the circuit (LOADED Q of the transformer primary).

In order to apply equation (35-2) one assumption must be made. That is, the unloaded Q (Q_c) of the transformer primary is 50 (this is not unreasonable in view of the use of a powdered iron core and other factors involved in the construction of a miniature IF transformer). Since the bandwidth (20 kc) and the resonant frequency (450 kc) of the example problem are known, the loaded Q (Q_T) of the circuit may be determined by rearranging the following equation and inserting known values.

$$BW = \frac{f_o}{Q} \qquad (11\text{-}20)$$

Solving for Q, which will be termed Q_T:

$$Q_T = \frac{f_o}{BW}$$

where: f_o = resonant frequency in cps
BW = bandwidth in cps
Q_T = previously defined

Inserting values from example:

$$Q_T = \frac{45 \times 10^4}{2 \times 10^4}$$

$$Q_T = 22.5$$

Inserting known and assumed values in equation (35-2):

$$X_{L_p} = \frac{R_o\,(Q_c - Q_T)}{2\,Q_c\,Q_T} \qquad (35\text{-}2)$$

$$X_{L_p} = \frac{40 \times 10^3\,(50 - 22.5)}{2 \times 50 \times 22.5}$$

$$X_{L_p} = \frac{11 \times 10^5}{2.25 \times 10^3}$$

$$X_{L_p} = 4.88 \times 10^2 \text{ or } 488 \text{ ohms}$$

Thus, the reactance of the transformer primary is 488 ohms.

The value of primary inductance (L_p) is found by rearranging the equation for X_L:

$$X_L = 2\pi f L \qquad (9\text{-}26)$$

Rearranging and inserting values:

$$L_p = \frac{X_{L_p}}{2\pi f}$$

$$L_p = \frac{488}{6.28 \times 45 \times 10^4}$$

$$L_p = 172.6 \times 10^{-6}$$

$$L_p = 172.6 \text{ microhenrys}$$

At resonance in a tuned circuit $X_L = X_C$. Therefore, the capacitor used to resonant the tuned circuit will also have a reactance of 488 ohms.

The value of total capacitance to be connected in parallel with L_p may be found by rearranging the equation for X_C:

$$X_C = \frac{1}{2\pi f C} \qquad (10\text{-}24)$$

Rearranging and inserting values:

$$C_T = \frac{1}{2\pi f X_C}$$

$$C_T = \frac{1}{6.28 \times 45 \times 10^4 \times 488}$$

$$C_T = 724 \text{ picofarads}$$

where: C_T = total capacitance needed for resonance (in farads).

X_C is in ohms and f is in cps.

Figure 35-5 shows a partial schematic of the interstage coupling network, including the tank capacitance (C_p), and the transistor output capacitance (C_{oe}). Notice that the output capacitance (C_{oe}) of transistor Q_1 appears in parallel with the tuned tank formed by C_p and L_p. Since

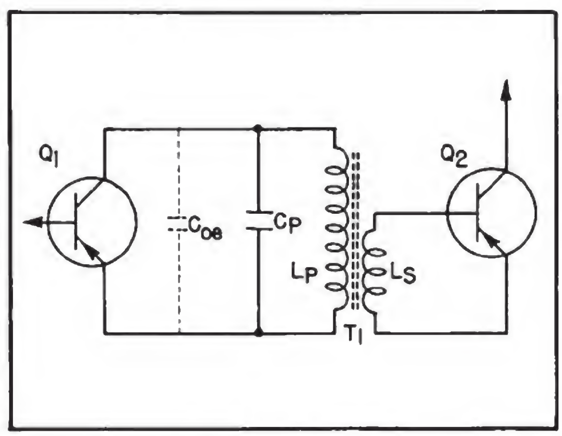

Figure 35-5 - Partial schematic of interstage coupling showing C_{oe}.

capacitors in parallel add, the value of capacitor C_p will be the value of total capacitance MINUS THE OUTPUT CAPACITANCE C_{oe}.

NOTE: For purposes of simplicity only the two major capacitances C_p and C_{oe} are being considered in this discussion. However, in the design of a practical circuit, there will be additional stray and reflected capacitances which appear in parallel with the primary and must be added to C_p and C_{oe} to obtain the total primary capacitance C_T. Usually these stray and re-

774-639 O - 65 - 8

flected capacitances are only a small portion of C_T and can, therefore, be neglected. If it is desired to include them, the value of C_{oe} may be increased by a small amount to represent their presence.

It was determined, by use of equation (10-24), that the value of total capacitance needed for resonance is 724 picofarads. Since a nominal value of C_{oe} is 40 picofarads, C_p will have a value of:

$$C_p = C_T - C_{oe}$$

$$C_p = 724 \times 10^{-12} - 40 \times 10^{-12}$$

$$C_p = 684 \text{ picofarads}$$

Although the interstage coupling network will function when connected as shown in Figure 35-5, the operation will not be as efficient as normally desired. The reason for this can be shown as follows. It has been stated that the selectivity of a superhet receiver is determined mainly by the selectivity of the IF stages. Since the selectivity of a circuit is primarily a function of circuit Q (Q_T), it is desirable to have a relatively high circuit Q. Q is a direct function of the inductive reactance, and therefore the inductance, present in the circuit. Thus, if the primary inductance (L_p) of the IF transformer could be increased WITHOUT AFFECTING THE TURNS RATIO PREVIOUSLY ESTABLISHED, the Q (and thereby the selectivity) of the circuit would be considerably improved. The inductance of the primary winding can be increased by increasing the number of its turns. If care is taken in the design and construction of the transformer, the number of turns can be increased without appreciably increasing the inherent resistance of the winding. However, to maintain the original turns ratio and impedance matching, the number of turns (hence the inductance) in parallel with Q_1's collector to emitter MUST REMAIN UNCHANGED. This is accomplished by TAPPING the winding as shown in Figure 35-6. Notice that the inductance of L_{p2} is the same value as L_{p1}. This indicates that the number of primary turns has been doubled. However, due to the action of mutual induction between the two sections, the inductance of the entire primary winding is NOT the algebraic sum of the two sections. The reason for this is shown by the following.

From chapter on inductance (Chapter 9):

$$L_T = L_1 + L_2 \pm 2M \qquad (9\text{-}9)$$

and:

$$M = k\sqrt{L_1 L_2} \qquad (9\text{-}7)$$

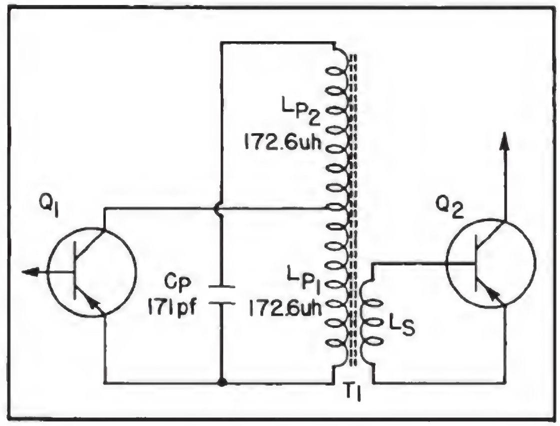

Figure 35-6 - Partial schematic of interstage coupling showing tapped primary.

Since k has been assumed to equal unity (due to the use of a powdered iron core), and L_1 is equal to L_2 it follows that, in this case:

$$M = 1\sqrt{(L_1)^2}$$

or:

$$M = L_1 \text{ (or } L_2\text{)}$$

Thus, since the two coils are connected series aiding, use equation (9-9) with L_1 in place of M.

$$L_t = L_1 + L_2 + 2L_1$$

or:

$$L_T = 4L_1$$

Thus, the total inductance of the primary winding in Figure 35-6 will be:

$$L_T = 4 \times 172.6 \times 10^{-6}$$

$$L_T = 690.4 \text{ microhenrys}$$

NOTE: The simplification, $L_T = 4L_1$, only applies in a situation where $k = 1$ and $L_1 = L_2$. If these conditions are not satisfied, then the complete equations (9-9) and (9-7) must be used.

In order to maintain the original operating frequency (450 kc), the LC product must remain unchanged. Since the total inductance of the primary has been increased to a value four times as large, then the total capacitance of the primary (C_T) must be decreased to a value one-fourth of its previous value. Therefore, the new value of C_T will be:

$$C_T = \frac{724 \times 10^{-12}}{4}$$

$$C_T = 181 \text{ picofarads}$$

The effect of the transistor output capacitance will also be reduced by a factor of four. Therefore, the new value of primary capacitance (C_p) will be:

$$C_p = C_T - \frac{C_{oe}}{4}$$

$$C_p = 181 \times 10^{-12} - \frac{40 \times 10^{-12}}{4}$$

$$C_p = 171 \text{ picofarads}$$

Figure 35-6 can be used to summarize the interstage coupling network. One function of the IF transformer is to provide impedance matching between the output impedance of one stage and the input impedance of the following stage. This is accomplished in Figure 35-6 by the 4.47:1 turns ratio between the L_{p1} primary section and the secondary winding L_s. Notice that the portion of the primary winding between the collector and emitter of Q_1 determines the turns ratio. The L_{p2} section of the primary does not enter into the calculation of turns ratio.

Another function of the IF transformer, and its associated tuned circuit, is to provide the selectivity for the receiver. It was shown how the relatively high Q, necessary for good selectivity, was obtained by increasing the inductance of the primary winding while, at the same time, maintaining the impedance match through tapping the primary.

An additional benefit gained is the reduction in the size of the capacitance used to resonate the transformer at the IF frequency. The reduction in value permits the physical size of the capacitor to be reduced to the point where it can be included within the metal shield in which the IF transformer is enclosed. In order to reduce undesirable feedback, the magnetic flux produced by the transformer windings must be prevented from inducing currents and voltages in nearby wires and components. This is accomplished by surrounding the transformer with a grounded metal shield or "can."

Q5. Two parallel resonant circuits are constructed. In one circuit $L_T = 250$ microhenrys and $C_T = 500$ picofarads. In the other circuit $L_T = 833$ microhenrys and $C_T = 150$ picofarads. Will the resonant frequency of the two circuits be approximately the same? Explain.

Q6. If the value of transistor output capacitance (C_{oe}) in the example problem of section 35-3 were increased, would the primary tank capacitance (C_p) have to be increased or decreased to maintain the same resonant frequency? Explain.

Q7. What would be the total inductance of a tapped primary winding (such as in Figure 35-6) if section $L_{p1} = 100$ microhenrys and section $L_{p2} = 300$ microhenrys? Assume $k = 1$.

BANDPASS CHARACTERISTICS OF IF AMPLIFIERS

As stated previously, the IF amplifier has the function of determining receiver selectivity and providing the major portion of the receivers radio frequency amplification. An additional function of the IF amplifier is to preserve all the original modulating intelligence by maintaining a sufficiently wide overall bandwidth.

Amplitude modulation of a carrier produces an upper and lower sideband frequency for every modulating frequency. These sideband frequencies are situated above and below the carrier, and are separated from the carrier by an amount equal to the original modulating frequency which produced them. Thus, if the highest audio frequency in the original modulating signal were 8 kc, then an upper sideband frequency of the carrier PLUS 8 kc, and a lower sideband frequency of the carrier frequency MINUS 8 kc would be present in the IF stage signal. Therefore, the overall bandwidth of the IF stage must be wide enough to pass these two sideband frequencies, and all sideband frequencies between them, with relatively equal amplification. For this particular application, the receiver would have to have an overall bandwidth of at least 16 kc. The overall bandwidth that a receiver is required to have is determined primarily by the job it is required to perform. For instance, a high fidelity receiver, which is required to reproduce the entire audio frequency range, would by necessity have a wider bandwidth than a military communications receiver which is required to reproduce only voice frequencies.

The purpose of the following sections is to illustrate how the overall bandwidth and response characteristics of the receiver may be altered by variations in the construction and tuning of the IF transformers. The various tuning methods to be discussed are SINGLE TUNING, DOUBLE TUNING, and STAGGER TUNING.

35-4. Resonant Circuit Recognition

Before entering into a discussion of IF tuning, a clarification of resonant circuit configurations will prove advantageous. In many cases a circuit that appears to be a parallel resonant circuit, will, in actuality, be a series resonant circuit.

Specifically, a transformer coupled IF stage is just such a case. Consider the simplified

A5. Yes. The LC product of the two circuits is approximately the same, therefore, the resonant frequency will be approximately the same.

A6. Decreased. $C_p = C_T - C_{oe}$.

A7. $L_T = 746.4$ microhenrys. The inductance values of the two sections are not equal, therefore, equations (9-9) and (9-7) must be used to obtain the solution.

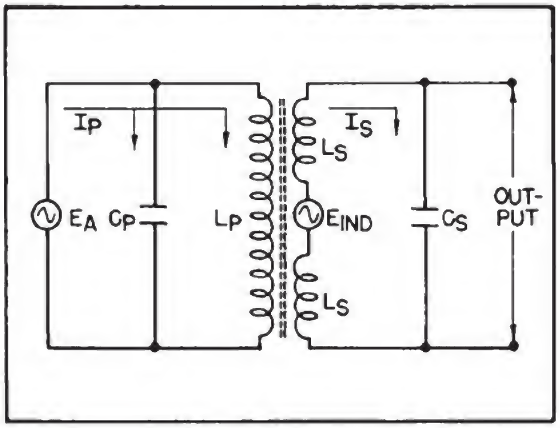

Figure 35-7 - Simplified transformer coupling of resonant circuits.

circuit of a transformer coupled IF amplifier in Figure 35-7.

Recognition of series or parallel configurations of tuned circuits is simplified, when an obviously defined source of voltage exists. The primary circuit of the transformer in Figure 35-7 has an obviously defined source of applied voltage (E_a). The voltage generator E_a represents the voltage produced by an amplifier (electron tube or transistor) of which the tuned circuit is the load. The tank circuit components (C_p) and (L_p) are in parallel with E_a. C_p and L_p provide a parallel path for the flow of primary current I_p. Therefore, the primary circuit of the transformer is a parallel resonant circuit because the voltage source is in parallel with the tank components.

When the voltage source of the circuit is not clearly defined, the recognition of circuit configuration is more difficult. Consider the secondary circuit of the transformer in Figure 35-7. At first glance it appears that the output is being taken across the parallel combination of L_s and C_s. However, when it is realized that the voltage source for the circuit is the induced voltage (represented by voltage generator E_{IND}) of the secondary winding, it can be seen that there is only one path for current flow (I_s) through C_s and L_s. The source voltage (E_{IND}) is effectively applied in series with the tank components. Therefore, the secondary circuit of the transformer is actually a series resonant circuit with the output being taken across the capacitor.

To simplify the discussion of transformer coupled tuned circuits, it is advantageous to have both the primary and secondary appear as the same configuration. Since any parallel circuit may be converted to an equivalent series circuit (section 12-7), the primary circuit is usually the one converted to an equivalent series circuit during the discussion of overall circuit response.

35-5. Single Tuned Transformer Coupled IF Stage

A SINGLE TUNED, transformer coupled, IF stage is one which contains only a single tuned resonant circuit. A single tuned IF amplifier stage was shown in Figure 35-1. Notice that the amplifier stage of Q_1 contains only one resonant circuit (C_5 and the primary of T_2).

It has been shown that when a capacitance and an inductance are connected to form a tuned resonant circuit, this circuit will exhibit frequency discriminatory properties. In other words, the circuit will favor the development of voltage or current at frequencies within its bandpass and discriminate against frequencies that fall outside its bandpass.

The behavior of a circuit, over a range of frequencies, is illustrated by a graph of voltage or current versus frequency, called a response curve. The response curve of a single tuned stage, illustrating the effect of circuit Q on voltage amplitude and bandwidth, is shown in Figure 35-8. Notice that the bandwidth BW_1

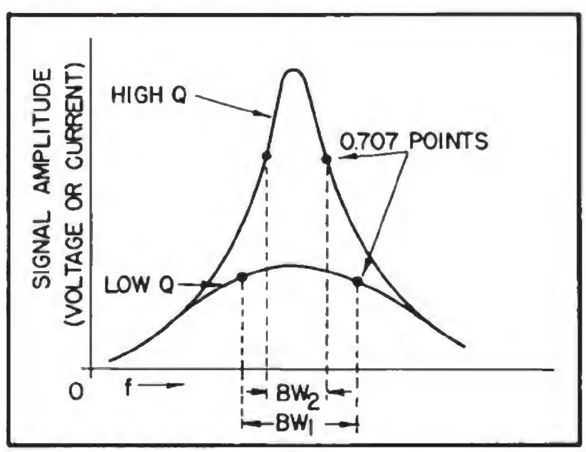

Figure 35-8 - Bandpass of single tuned high Q and low Q circuits.

(between the half-power points) of the low Q circuit is much wider than the bandwidth BW_2 of the high Q circuit. Due to the slope of the sides of the response curves, frequencies falling outside the bandpass of the high Q circuit will receive greater attenuation with respect to f_o than will frequencies falling outside the bandpass of the low Q circuit. Thus, a receiver employing high Q circuits will have good selectivity. The method for achieving high Q tuned circuits, and thereby good selectivity in a transistor receiver, was explained in section 35-3. The high Q curve in Figure 35-8 represents the response of only a single IF stage containing a single tuned circuit. It was shown (section 32-5) that when two or more amplifier stages or tuned circuits, all tuned to the same center frequency, are connected in cascade (following one another), the overall bandwidth will be less than the bandwidth of any individual tuned circuit. Therefore, while the Q of the IF stages tuned circuits should be high, to provide good selectivity, they must not be so high as to reduce the overall bandwidth to the point where the sideband frequencies representing the highest modulating frequencies are attenuated.

Use of the single tuned stage yields very satisfactory results in many applications. High gain is easily obtained. It is evident from the high peak at resonance and the steep slope of the curve for a high Q circuit (Figure 35-8) that selectivity is good. However, it is also evident (from a comparison of the amplitude of secondary voltage at the resonant peak to the amplitude of voltage at the 0.707 points) that all frequencies within the bandpass will not receive equal amplification. The sideband frequencies, representing the low modulating frequencies (frequencies closest to f_o), will be amplified to a greater extent than the sidebands representing the high modulating frequencies.

35-6. Double Tuning of Transformer Coupled IF Stage

In some applications (such as high fidelity receivers) the bandwidth requirements are such that a single tuned stage will not be able to meet them. A specific application may require a wider bandwidth than is possible with single tuning. Another application may require that all sideband frequencies within the bandpass receive relatively equal amplification. Still another application may require a wide bandpass but also a high degree of selectivity for frequencies immediately outside the bandpass. These requirements can be satisfied by the use of DOUBLE TUNING in the IF stages. Double tuning refers to an interstage transformer in which BOTH the primary and the secondary contain resonant circuits.

The bandpass characteristics of a double tuned stage depend on many things, among which are the coefficient of coupling (k) between the primary and secondary windings, the Q's of the primary and secondary circuits, and the mutual inductance.

An important property of inductively coupled (transformer) circuits is mutual inductance (M). Mutual inductance (as explained in section 9-16) is the common property of inductively coupled circuits that determines, for a particular frequency and current in one circuit, the amount of mutually induced voltage in the other circuit. The following relation exists between the current and induced voltage in an inductively coupled circuit.

$$E_S = 2\pi f M I_P \qquad (35\text{-}3)$$

where: E_S = secondary voltage in volts
f = frequency in cps
M = mutual inductance in henrys
I_P = primary current in amps

This relation is similar to that for self-induced voltage (E_{IND}). The equation for E_{IND} is derived from Ohm's law for ac circuits:

$$I = \frac{E}{X_L} \qquad (9\text{-}27)$$

Transposing for E:

$$E = X_L \times I$$

Since: $X_L = 2\pi f L$

Substituting:

$$E_{IND} = 2\pi f L I \qquad (35\text{-}4)$$

where: E_{IND} = self-induced voltage in volts
f = frequency in cps
L = inductance in henrys
I = current in amps

Although problems involving inductively coupled circuits are very complex, they may be simplified to a certain extent if the following assumptions are made:

1. When a secondary circuit, containing an impedance (Z_S), is coupled to a primary circuit the effect on the primary is the same as if an equivalent impedance (Z_C), herein after referred to as the coupled impedance, had been connected in series with the primary. The equation for the coupled impedance (derived at the end of

this chapter) is:

$$Z_C = \frac{(2\pi f M)^2}{Z_S} \qquad (35\text{-}5)$$

where: Z_C = coupled impedance (a vector quantity).
f = frequency in cps
M = mutual inductance in henrys
Z_S = secondary impedance (a vector quantity).

2. The secondary voltage (E_S), induced by the primary current (I_P), has the value $2\pi f M I_P$ from equation (35-3) and lags the primary current by 90°.

3. The secondary current (I_S) is that value of current that would flow in the secondary if the primary were removed and the induced secondary voltage (E_S) were applied in series with the secondary coil.

The coupled impedance is a vector quantity of the form R plus or minus j X. The primary impedance is also a vector quantity of the form R_P plus or minus jX_P. Z_C has the same phase angle as Z_S but is of the opposite sign. This means that when Z_C is assumed to be reflected in series with the primary, capacitive reactance in the secondary will be reflected as inductive reactance in the primary, and if the secondary exhibits inductive reactance it will be reflected to the primary as capacitive reactance. The total impedance of the primary (Z_P') will be the vector sum of the coupled impedance Z_C and the impedance of the primary circuit Z_P considered by itself.

It has been shown that mutual inductance is a direct function of the coefficient of coupling (k). In other words, if the coefficient of coupling is very low, the mutual inductance will be low, the secondary impedance will have very little effect on the primary because Z_C will be small (small value of M in equation 35-5), and the secondary induced voltage will be small (small value of M in equation 35-3). Therefore, a very small value of k will cause the primary and secondary circuits to behave essentially like two separate circuits.

Increasing the coefficient of coupling will increase the value of mutual inductance and thereby, increase the value of coupled impedance in the primary and, UP TO A POINT (explained shortly), the induced secondary voltage will be increased.

It has been stated that the bandwidth characteristics of a double tuned, transformer coupled, stage depend on part on the amount of coupling. This can be shown by use of the equivalent circuit and response curves in Figure 35-9. Notice that the coupled impedance (Z_C) is represented as being in series with the primary circuit. Applied voltage (E_a) and internal resistance (r) represent the amplifier (tube or transistor) voltage and resistance. R_{Peq} is the series equivalent of the original parallel circuit resistance. R_{Seq} is the equivalent resistance of the secondary circuit. It will be assumed, for purposes of explanation, that the primary and secondary circuits have identical Q's, both circuits are tuned to the same resonant frequency, and the amount of coupling between the circuits is variable (by physically moving the coils). It will also be assumed that the Q of both circuits is held constant for all values of coupling.

To demonstrate the effect of coupling variations, the circuits are first moved far enough apart so that very little coupling takes place (low value of k). This condition is called LOOSE COUPLING. With loose coupling, there is very little transfer of energy between the primary and secondary circuits. The value of induced secondary voltage (E_S) is small because the value of mutual inductance (M) is small. The low value of M also causes the value of Z_C to be small (according to equation 35-5). The small value of Z_C has very little effect on the operation of the primary circuit. Due to the small amount of interaction between the primary and secondary, the two circuits behave essentially as if they were separate tuned circuits. Except for the slope of the sides being slightly steeper, the response curve for loose coupling in Figure 35-9 is the same as for a single tuned circuit.

As the coefficient of coupling is increased (the circuits are continually moved closer together), the value of mutual inductance increases. This continually increases the amount of voltage induced in the secondary winding. The value of the coupled impedance is also increasing, and due to its effect on the primary current, the response curve becomes wider (bandwidth increases).

As the coefficient of coupling is continually increased Z_C, which is resistive at the resonant frequency, continues to increase until eventually a value of coupling is reached where Z_C equals the equivalent resistance of the primary (R_{Peq}). This condition is called CRITICAL COUPLING (k_c) and since at this point the reflected resistance matches the primary resistance there is a maximum transfer of energy between the circuits. Thus, the induced secondary voltage is at its maximum value, as shown by the response curve for critical coupling in Figure 35-9. Notice that the critical coupling curve is flattened slightly on top. At resonance both primary and secondary circuits appear resist-

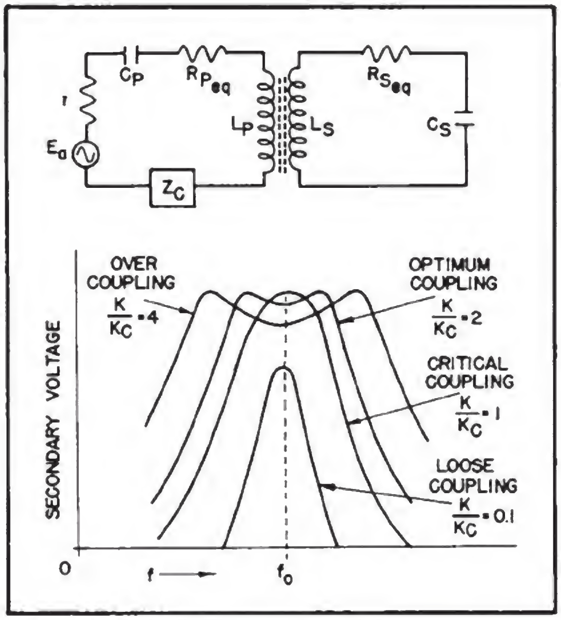

Figure 35-9 - Equivalent double tuned circuit and response curves.

ive. However, at frequencies above resonance, a series circuit appears inductive, and at frequencies below resonance a series circuit appears capacitive. It was stated that if the secondary circuit appeared inductive the impedance coupled into the primary would have the same phase angle but would appear capacitive in nature. In a single tuned circuit at frequencies slightly off resonance, the primary current begins to decrease, however, a different situation exists in double tuned circuits. At frequencies above resonance, the primary circuit (considered here as an equivalent series circuit) takes on inductive characteristics, however, the capacitive portion of the complex coupled impedance (Z_C) will cancel a portion of the primary inductance. This will, in effect, reresonate the circuit at a slightly higher frequency, or it could be said to extend the resonant condition of the circuit slightly beyond f_o. A similar action takes place below resonance, where the primary circuit begins to appear capacitive, the coupled impedance appears inductive in nature and cancels some of the primary capacitance. It could also be said that this action extends the condition of resonance slightly below f_o. In any case, the result is to cause the primary current to remain high within most of the bandwidth (as shown by the flat top of the critical coupling response curve).

Critical coupling (k_C) has been defined as that value of coupling (k) at which maximum transfer of energy between primary and secondary is achieved. Critical coupling is determined, mathematically, by the following equation (the derivation of which is not presented):

$$k_c = \frac{1}{\sqrt{Q_P \times Q_S}} \qquad (35\text{-}6)$$

where: k_c = critical coupling (a decimal number between zero and one).

Q_P and Q_S = primary and secondary circuit Q respectively.

When $Q = Q_P = Q_S$ the above equation may be reduced to:

$$k_c = \frac{1}{Q} \qquad (35\text{-}7)$$

Use of the above equations in determining the critical coupling factor of a circuit is shown by the following example. Assume that the value of $Q = Q_P = Q_S$ for the circuit in Figure 35-9 is 50. Since the primary and secondary circuit Q's are equal, equation (35-7) may be used (if Q's were NOT equal, equation (35-6) would be used).

$$k_c = \frac{1}{Q} \qquad (35\text{-}7)$$

$$k_c = \frac{1}{50}$$

$$k_c = 0.02$$

When the value of coupling k equals the value of critical coupling k_c, the ratio of k to k_c will be one:

$$\text{when } k = k_c\text{: } \frac{k}{k_c} = 1$$

The response curves of Figure 35-9 are representative of any double tuned circuit (assuming Q_P and Q_S equal 25 or greater). Therefore, determining the value of k_c, and then solving the given ratios of k/k_c for k, will permit determination of the value of coupling needed in a circuit to obtain a response curve similar to those shown.

For instance, using the value of $k_c = 0.02$ determined above, it is desired to find the value of k which will give the loose coupling response curve shown in Figure 35-9. Transpose the ratio $k/k_c = 0.1$ for k:

$$k = 0.1 \times k_c$$

Insert known values and solve:

$$k = 0.1 \times 0.02$$

$$k = 0.002$$

Thus, a coefficient of coupling for this particular circuit (assumed Q of 50) of 0.002 will give a single peaked resonance curve of the relative amplitude and bandwidth of the loose coupling curve. For values of k between 0.002 and 0.02, the amplitude and bandwidth will continue to increase until the critical coupling curve is obtained. Changing the value of primary and secondary circuit Q will cause the value of k and k_c to change but the appearance of the curve for a specific ratio will remain the same.

Q8. Determine k and k_c for a circuit where $Q_P = Q_S = 80$ when a response curve similar to the loose coupling curve ($k/k_c = 0.1$) is desired.

When the coefficient of coupling is increased beyond the value of k_c, the value of primary current at resonance begins to decrease. This is caused by the continued increase of the coupled impedance (which at resonance is resistive) and is indicated by the dip in the response curve, at f_o, of the OPTIMUM COUPLING curve in Figure 35-9. It can be seen that the points above and below f_o, at which the primary circuit reactances and the reflected reactances of Z_C reresonate, have moved further apart. The value of k for which optimum coupling is achieved is twice the value of k_c. Therefore, the coefficient of coupling (in the circuit where $k_c = 0.02$) needed to obtain a curve similar to the optimum coupling curve is determined by use of the k/k_c ratio:

$$\frac{k}{k_c} = 2$$

Transpose and insert values

$$k = 2 \times 0.02$$

$$k = 0.04$$

A continued increase in the coefficient of coupling will result in further reduction of the gain at resonance and a further increase in the distance between the two resonant peaks (above and below f_o). The circuits are now said to be OVER COUPLED and the overall response of the stage will take on the appearance of the double humped curve in Figure 35-9.

In electron tube IF stages, where the plate and grid resistances are very high, the voltage gain of one double tuned stage for the condition of critical coupling (assuming $Q_P = Q_S$) is approximately half of that for one single tuned stage. In transistor IF stages (where the output and input resistances are relatively low) this approximation will not always be true. However, it can be said that double tuned stages will yield less voltage gain per stage than single tuned stages (per stage) PROVIDING THE BANDPASS OF THE TWO STAGES ARE THE SAME. Comparing two cascaded single tuned stages to two cascaded double tuned stages, the single tuned stages will have only slightly more gain per stage than the double tuned stages. As additional stages are added, the double tuned stages will have more voltage gain PER STAGE than the single tuned stages. The reason for this is seen in the qualification that the overall bandpass of the two types of circuits MUST BE THE SAME. The only way the bandpass of the single tuned stages can be increased, in order to match that of the double tuning, is to decrease the Q of the single tuned circuits, which will, of course, decrease the gain per stage of the single tuned circuits.

For double tuned circuits, the bandwidth is determined primarily by the value of coupling, while the depth of the valley between the resonant peaks (and hence the uniformity of response) is determined by the relation of Q to k. Thus, if the circuit is operated with a value of coupling greater than critical, and k is held constant, a large Q means a deeper valley and a small Q means less of a valley.

From the foregoing it can be seen that the desired response curve is a compromise and is achieved by the proper manipulation of BOTH k and Q because they are interrelated.

Q9. An optimum coupling response curve is desired for a double tuned circuit where $Q_P = 90$ and $Q_S = 40$. Determine the approximate value of coupling (k) necessary.

35-7. Stagger Tuned IF Stages

In the previous sections, two methods of tuning the IF stages were discussed. It was shown that single tuning of the IF stages provided large voltage gain per stage. However, it was also stated that the large voltage gain of single tuning is possible only with a relatively narrow bandwidth and relatively high Q's. A wide bandpass can only be obtained, using single tuning, with a sacrifice of selectivity and gain. It was also pointed out that with single tuning the frequency response within the bandpass is not uniform.

It was then shown that it is possible to obtain a wide bandpass, while maintaining good selectivity, and an improvement in uniformity of response within the bandpass, by use of double

tuning. However, double tuning does not provide the voltage gain that the narrow band single tuning does.

A third type of tuning, called STAGGER TUNING, is available which combines the gain of the single tuned stage with the wide, uniform, bandpass of the double tuned stage. The resonant circuits in stagger tuned, transformer coupled, IF stages are isolated from each other so that no mutual coupling takes place between them. The isolation is necessary since each resonant circuit must have the characteristics of a single tuned circuit.

In previous types of tuning the resonant circuits were all tuned to the same frequency. In stagger tuning two transformer coupled IF stages, the two tuned circuits (one in each stage) are each tuned to a different resonant frequency, one slightly above the receiver IF frequency and the other slightly below the receiver IF frequency.

Figure 35-10 shows a simplified diagram (power supplies, biasing networks, neutralizing capacitors, etc. omitted) of a two stage, transformer coupled, stagger tuned IF section of a receiver. When tuned in this manner the stages are often called a STAGGERED PAIR.

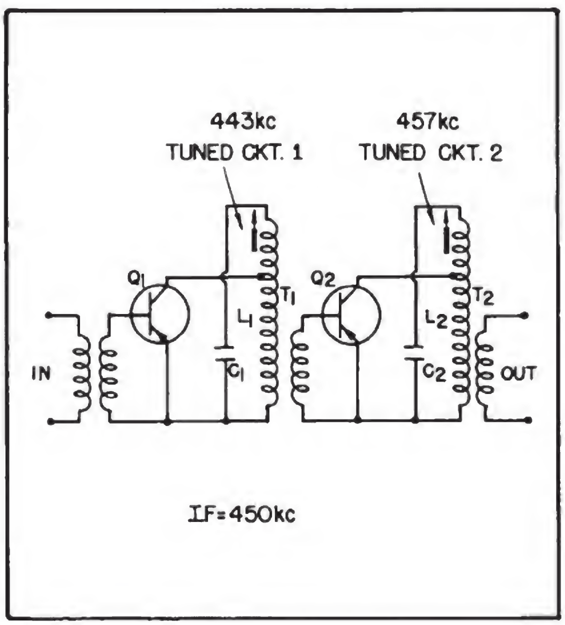

Figure 35-10 - Staggered pair IF amplifiers (simplified schematic).

Notice that, although the IF frequency of the receiver is 450 kc, tuned circuit number one (L_1 and C_1) is tuned to a resonant frequency of 443 kc, while tuned circuit number two (L_2 and C_2) is tuned to a resonant frequency of 457 kc. In the following discussion it is assumed that the frequency response and amplification of transistors Q_1 and Q_2 are identical, although in actuality there will always be some differences due to manufacturing tolerances. It will also be assumed that the Q and frequency response of the two tuned circuits are identical. Since there is no inductive coupling between tuned circuit number one and number two, each circuit will have the frequency response and characteristics of a single tuned circuit.

Figure 35-11A illustrates the individual response curves of the two tuned circuits. Circuit number one is tuned below the IF frequency of the receiver by 7 kc and has its resonant peak at 443 kc. The circuit is so designed to have a bandwidth (BW_1) of 14 kc. This is evident by the fact that the 0. 707 (half-power) points occur at 436 kc and 450 kc respectively. It is significant that the UPPER half-power point of tuned circuit number one occurs at the receiver IF frequency. The relative gain of circuit #1 at its resonant peak (f_{o1}) of 443 kc is approximately 140 (139. 69). The relative gain at the half-power points is very close to 100 (98. 69). Figure 35-11A shows tuned circuit #2 as having its resonant peak (f_{o2}) at 457 kc. The LOWER half-power point of circuit #2 occurs at the receiver IF frequency and its upper half-power point occurs at 464 kc. The bandwidth (BW_2) is also 14 kc. It might appear, at first glance, that the overall bandwidth of the staggered pair will be BW_1 plus BW_2, or 28 kc, however, this is not the case (as will be shown shortly). The amount of amplification, or gain, that a signal receives in passing through a number of stages is the product of the gain of the individual stages. One method of obtaining an overall response curve of two or more amplifier stages in series is to apply a number of individual frequencies to the input of the cascaded stages and plot the relative amplitude of the output for each frequency. When all these plotted points are connected together the resultant will be the overall response curve of gain versus frequency. If the relative gains of the individual stages, over a band of frequencies, are known, the overall response curve may be plotted directly from this data. Thus, the overall response curve of the staggered pair of IF amplifiers (Figure 35-10) may be plotted from the individual response curves of Figure 35-11A. For example, the upper half-power point of circuit #1 and the lower half-power point of circuit #2 coincide at the IF frequency of 450 kc. Since the gain of each circuit at the 0. 707 point is slightly less than 100, the overall gain of the staggered pair at the receiver IF frequency of 450 kc will be slightly less than $100 \times 100 = 10,000$ (actually 9739. 7).

A8. $k_c = 0.0125$; $k = 0.00125$.

A9. By equation (35-6) k_c equals approximately 0.0167. For optimum response $k/k_c = 2$. Therefore, k equals approximately 0.0334.

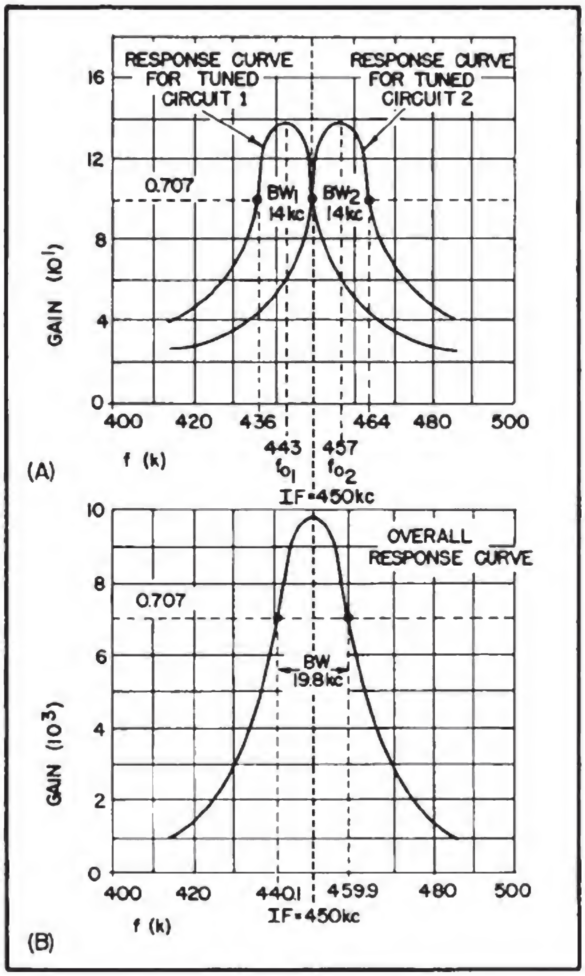

Figure 35-11 - Response curves of staggered pair IF amplifiers.

This point is shown as the resonant peak of the overall response curve in Figure 35-11B. Going above the receiver IF frequency to the resonant frequency of circuit #2 (457 kc) it is seen that the gain of circuit #1 has decreased considerably (down to 60) while the gain of circuit #2 is at its maximum point. The overall gain at 457 kc will be approximately $140 \times 60 = 8,400$ or (8.4×10^3). As the frequency is increased past this point it can be seen that the gain of both circuits is decreasing. When the upper half-power point of circuit #2 (464 kc) is reached the overall gain of the staggered pair will be approximately $100 \times 42.5 = 4.25 \times 10^3$, which is well below the half-power point of the overall response curve.

Points for the overall response curve below the receiver IF frequency are plotted in the same manner. Notice, that for a short distance above and below 450 kc the increasing gain of one tuned circuit and the decreasing gain of the other tuned circuit cause the overall response curve to maintain a relatively flat top. The rapidly decreasing gain of BOTH circuits after the individual resonant peaks have been passed cause the bandwidth of the overall response curve to be less than the sum of the individual bandwidths.

When the upper half-power point of one tuned circuit coincides with the lower half-power point of the other tuned circuit, in a staggered pair, the overall response curve will be similar to the critical coupling response curve of a double tuned stage, and the overall bandwidth will be 1.414 TIMES THE BANDWIDTH OF ONE OF THE TUNED CIRCUITS. Therefore, as seen in Figure 35-11B, the bandwidth of the staggered pair (when the half-power points overlap) is $1.414 \times 14 \times 10^3 = 19,796 \times 10^3$, or approximately 19.8 kc.

Varying the resonant frequencies of the two tuned circuits, in respect to the receiver IF frequency, will vary the overall response curve and bandwidth.

Q10. If the individual tuned circuits of a staggered pair were tuned further above and below the receiver IF frequency than in Figure 35-11A, would the overall gain of the pair increase or decrease?

35-8. Output Coupling

As stated previously, the output of a transistor IF amplifier may be transformer coupled to either another IF amplifier stage or to a detector stage. Although the method of coupling is the same, the characteristics of an IF transformer feeding a detector stage are slightly different than those of a transformer feeding another IF stage. The main difference being the slight variation in turns ratio, coefficient of coupling, etc., to facilitate impedance matching. This is caused by the fact that the input impedance of the detector stage is usually lower than that of an IF amplifier.

35-9. Derivation of Primary Reactance Equation

The derivation of equation (35-2), used in section 35-3, is hereby given.

The first step in deriving equation (35-2) is to change the partial schematic of the interstage coupling network, Figure 35-12, into an equivalent circuit.

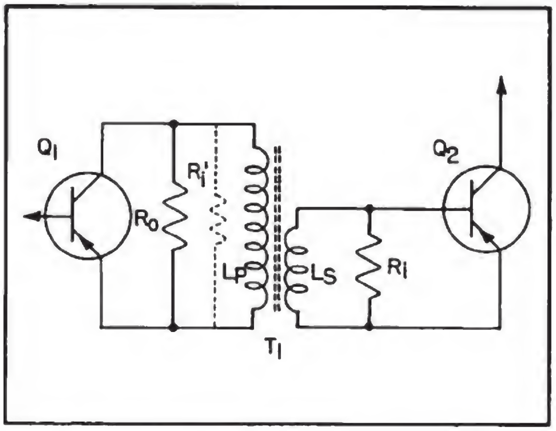

Figure 35-12 - Partial schematic of interstage coupling network.

NOTE: The symbols used in Figure 35-12 are explained in section 35-3 (in relation to Figure 35-4).

Figure 35-13 is the constant current equivalent of the interstage coupling network as seen by the collector of Q_1.

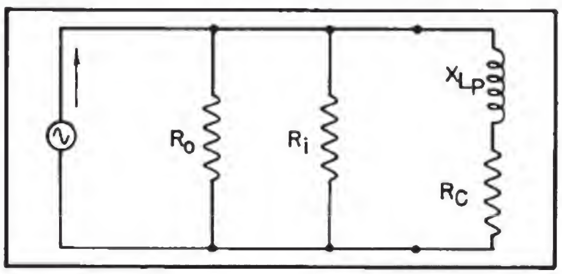

Figure 35-13 - Constant current equivalent of interstage coupling network.

The equivalent resistance of R_o and R'_i may be found by the product over the sum method:

$$R_{eq} = \frac{R_o R'_i}{R_o + R'_i}$$

However, for MAXIMUM transfer of power, R_o must equal R'_i. Therefore, when $R_o = R'_i$, the equivalent resistance, is half of either one:

$$R_{eq} = \frac{R_o}{2} = \frac{R'_i}{2}$$

Thus, the constant current equivalent circuit is simplified by substituting $R_o/2$ for the parallel combination of R_o and R'_i, and becomes Figure 35-14A.

Since a constant current equivalent circuit is being used, it is more convenient to consider the inductor as a parallel RL circuit than as a series RL circuit. Thus, branch B of Figure 35-14A is replaced with a parallel RL circuit, wherein X_{LP} represents the inductive reactance of the coil and R_C represents the losses of the coil. The equivalent circuit now appears as in Figure 35-14B. In this equivalent circuit, $R_o/2$

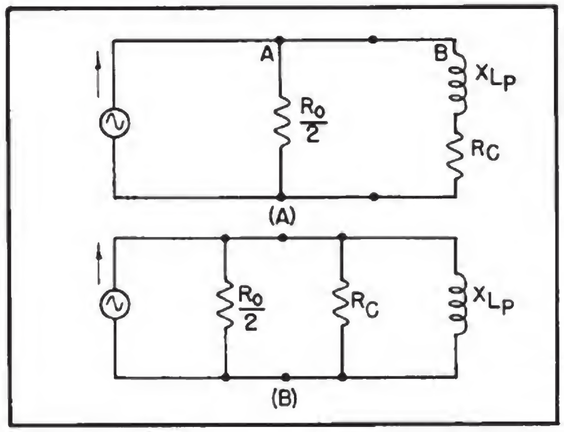

Figure 35-14 - Simplified constant current equivalent circuit.

represents the transistor Q_1's output resistance in parallel with the reflected input resistance of transistor Q_2. The primary winding of the interstage transformer is represented by the parallel combination of R_C and X_{LP}.

The Q of the coil alone (Q_C) is determined as R_C/X_L (since the resistance of the coil is considered to be in parallel with the coil. As an equation:

$$Q_C = \frac{R_C}{X_L}$$

The Q of the entire circuit must be computed using $R_o/2$ as well as R_C, since, in the complete circuit, both of these resistances are in parallel with X_{LP}. By the product over the sum method, the total resistance (R_T) of the equivalent circuit (Figure 35-14B) is:

$$R_T = \frac{\left(\frac{R_o}{2}\right) R_C}{\left(\frac{R_o}{2}\right) + R_C}$$

A10. Decrease. Because the product of the individual gains at any frequency would be less.

Simplifying the numerator:

$$R_T = \frac{\frac{R_o\, R_C}{2}}{\left(\frac{R_o}{2}\right) + R_C}$$

Simplify the denominator by finding the LCD:

$$R_T = \frac{\frac{R_o\, R_C}{2}}{\frac{R_o + 2\, R_C}{2}}$$

Inverting the denominator and multiplying:

$$R_T = \frac{R_o\, R_C}{R_o + 2\, R_C} \qquad (35\text{-}8)$$

The total Q of the entire circuit in equation form is:

$$Q_T = \frac{R_T}{X_L} \qquad (35\text{-}9)$$

Substituting equation (35-8) into equation (35-9) in place of R_T yields:

$$Q_T = \frac{\frac{R_o\, R_C}{R_o + 2\, R_C}}{X_L}$$

Inverting the denominator and multiplying:

$$Q_T = \frac{R_o\, R_C}{R_o + 2\, R_C} \times \frac{1}{X_L} \qquad (35\text{-}10)$$

It was shown previously that:

$$Q_C = \frac{R_C}{X_L}$$

Transposing for R_C yields:

$$R_C = Q_C\, X_L \qquad (35\text{-}11)$$

Substituting equation (35-11) into equation (35-10) in place of RC yields:

$$Q_T = \frac{R_o\, Q_C\, X_L}{R_o + 2\, Q_C\, X_L} \times \frac{1}{X_L}$$

Cancelling X_L's:

$$Q_T = \frac{R_o\, Q_C}{R_o + 2\, Q_C\, X_L}$$

It is desired to solve the above equation for X_L. First, multiply both sides by $(R_o + 2\, X_C\, X_L)$:

$$Q_T\, (R_o + 2\, Q_C\, X_L) = R_o\, Q_C$$

Expanding the left side:

$$Q_T\, R_o + 2\, Q_T\, Q_C\, X_L = R_o\, Q_C$$

Subtracting $Q_T\, R_o$ from both sides:

$$2\, Q_T\, Q_C\, X_L = R_o\, Q_C - Q_T\, R_o$$

Dividing both sides by $2\, Q_T\, Q_C$:

$$X_L = \frac{R_o\, Q_C - Q_T\, R_o}{2\, Q_T\, Q_C}$$

Factoring R_o in the numerator yields:

$$X_L = \frac{R_o\, (Q_C - Q_T)}{2\, Q_T\, Q_C}$$

Considering X_L as the inductive reactance of the primary and designating it X_{LP}, yields equation (35-2) as used in section 35-3.

$$X_{LP} = \frac{R_o\, (Q_C - Q_T)}{2\, Q_T\, Q_C} \qquad (35\text{-}2)$$

The great advantage of equation (32-2) lies in the fact that the required reactance for the tank inductance can be determined through use of a few easily acquired quantities, specifically, the transistor output resistance (R_o), the Q of the primary winding (Q_C), and the desired overall circuit Q (Q_T).

35-10. Derivation of Coupled Impedance Equation

This section is devoted to the derivation of the equation (35-5) for coupled impedance (Z_C).

$$Z_C = \frac{(2\pi\, f\, M)^2}{Z_S} \qquad (35\text{-}5)$$

Equation (35-5) is derived from the relation that (in a mutually coupled circuit) a voltage, E_{MP}, is induced in the primary by the secondary current I_S (just as a voltage is induced in the secondary by the primary current). The RATIO of the voltage (E_{MP}) induced in the primary (by the presence of the secondary) to the primary current (I_P) is called the coupled im-

pedance (Z_C). In equation form:

$$Z_C = \frac{E_{MP}}{I_P} \qquad (35\text{-}12)$$

The mutual voltage (E_{MP}) induced in the primary by the secondary is equal to the product of the secondary current (I_S), the mutual inductance (M), the frequency (f), and the constant 2π. In equation form:

$$E_{MP} = 2\pi f M I_S \qquad (35\text{-}13)$$

Substituting equation (35-13) into equation (35-12) yields:

$$Z_C = \frac{2\pi f M I_S}{I_P} \qquad (35\text{-}14)$$

Applying Ohm's law to the secondary:

$$I_S = \frac{E_S}{Z_S} \qquad (35\text{-}15)$$

Substituting (35-15) for the secondary current in equation (35-14) yields:

$$Z_C = \frac{2\pi f M \frac{E_S}{Z_S}}{I_P} \qquad (35\text{-}16)$$

The mutually induced secondary voltage (E_S) is equal to the product of the primary current (I_P), the mutual inductance (M), the frequency (f), and the constant 2π. This was introduced in section 35-6 as equation (35-3):

$$E_S = 2\pi f M I_P \qquad (35\text{-}3)$$

Substituting equation (35-3) in equation (35-16) yields:

$$Z_C = \frac{2\pi f M \frac{2\pi f M I_P}{Z_S}}{I_P}$$

Inverting the denominator yields:

$$Z_C = \frac{(2\pi f M)\,(2\pi f M I_P)}{Z_S} \times \frac{1}{I_P}$$

Multiplying:

$$Z_C = \frac{(2\pi f M)^2\, I_P}{Z_S I_P}$$

Cancelling I_P yields equation (35-5) as used in section 35-6.

$$Z_C = \frac{(2\pi f M)^2}{Z_S} \qquad (35\text{-}5)$$

EXERCISE 35

1. Is the value of intermediate frequency the same in every receiver? Explain.
2. Explain the main disadvantages of a low value of intermediate frequency.
3. Why are the CB and CC configurations seldom used as IF amplifiers in a transistor receiver?
4. What are the advantages of using more than one stage of IF amplification?
5. Is it possible for the LC product of a 200 kc resonant circuit to be the same as the LC product of a 300 kc resonant circuit? Explain.
6. What is meant by permeability tuning?
7. Explain the difference between the loaded and unloaded Q of an IF transformer.
8. Determine the turns ratio of a transformer used to match the 4 k ohm input resistance of a transistor to the 55 k ohm output resistance of another transistor.
9. What is the necessary value of input resistance (R_i) for maximum power transfer if a transformer having a turns ratio of 5:1 is used with a transistor having an output resistance of 50 k ohms?
10. What is the value of primary reactance in a transformer coupled, 465 kc, IF stage when R_o = 30 k ohms, BW = 30 kc and the unloaded Q of the transformer is 70?
11. What is the value of primary inductance for the IF stage of question 10?
12. What value of total capacitance (C_T) would be needed to resonate the IF stage of question 10?
13. Assume the transistor output capacitance (C_{oe}) in question 10 is 44 picofarads. What value of primary capacitance (C_P) would be connected across the transformer primary?
14. Explain why the total inductance of a tapped coil is not the algebraic sum of the individual sections.
15. Explain the reasons for, and the advantages of, using a tapped transformer in the IF section of a transistor receiver.
16. Explain why the selectivity of a high Q circuit is better than that of a low Q circuit.
17. What are some disadvantages of an extremely high Q circuit when used in an IF stage?
18. Explain the term "coupled impedance."
19. Explain what is meant by the term "loose coupling."
20. Why do circuits that are loose coupled behave essentially as separate tuned circuits?
21. Define critical coupling.
22. If high Q circuits are used will the value of critical coupling (k_c) be large or small? Explain.
23. Explain the reason for the flattening of the top of the critical coupling response curve.
24. Explain the reason for the slight dip in the optimum coupling response curve.
25. Explain over coupling. Will the ratio k/k_c be more than one or less than one when circuits are over coupled?
26. What primarily determines the depth of the valley in a double tuned, over coupled, circuit.
27. What is meant by stagger tuning?
28. What are the main advantages of staggered pairs?
29. Assume the upper half-power point of one circuit and the lower half-power point of another, in a staggered pair, overlap. Why is the voltage gain within the overall bandpass GREATER than that of either circuit alone?
30. Why would the overall bandpass of the staggered pair in question 29 be LESS than the sum of the bandpass of the two circuits?

CHAPTER 36

SOLID STATE DETECTOR

DETECTION, or DEMODULATION, is the process of recreating the original modulating frequencies (intelligence) from the radio frequencies present in the composite IF signal.

When a radio signal is amplitude modulated, the intelligence is contained in the relationship between the carrier and sideband frequencies. Applying the composite IF signal to a circuit containing a non-linear conducting device (or non-linear impedance) allows recreation of the original modulating signal. The circuit, in which this recreation is accomplished, is called the DETECTOR or DEMODULATOR.

The term demodulator is used because the process of detection is considered to be the opposite of modulation. The non-linear conducting device used in solid state detector circuits may be either a PN junction diode or the input junction of a transistor. The IF signal consists of many radio frequencies which continuously shift in phase with respect to each other. Graphing the instantaneous sums of these frequencies with respect to time will result in a symmetrical radio frequency envelope. The amplitude variations in the positive half of this envelope are equal and opposite to the amplitude variations in the negative half of the envelope. Thus, the average value of the composite IF signal is zero. As long as the average value of the signal is zero no useful energy may be extracted. The non-linear characteristics of the detector circuit will cause the IF signal to be distorted, the positive half cycles will be different than the negative half cycles (or vice-versa depending on the connection of the diode). The distortion introduced by the detector circuit will cause the average value of the waveform to be other than zero and vary in accordance with the original modulating signal. The output of the detector (before filtering) will contain, in addition to the carrier and sideband frequencies, the original modulation frequencies.

This chapter will discuss the theory and operation of series and shunt detectors using a PN junction diode as the non-linear device. Next a detector circuit utilizing a transistor as the non-linear device will be discussed. Finally, methods of output coupling and obtaining an automatic volume control voltage will be analyzed.

36-1. Solid State Diode Detectors

Broadcast band receivers, along with many other types of receivers employing the superheterodyne principle of reception, make extensive use of solid state diodes in the detector stage. The reason for their popularity stems from the fact that they are; space saving due to their small size, require no power supply for operation, exhibit a low internal capacitance and low dynamic resistance.

Diode detector circuits produce an output voltage and current which varies in accordance with the original modulating signal.

The symmetrical radio frequency waveform, which is the output of the IF stage, is the input waveform to the detector stage.

36-2. Series Diode Detector

The circuit in Figure 36-1 shows the output of the IF amplifier being inductively coupled to the input of the SERIES DETECTOR stage. This circuit is called a series detector because the diode is in series with the load resistor R_L. The voltage which is developed across the secondary tuned circuit of T_1 is now the source for the detector. When the resultant RF variations of the input signal cause the top of the secondary tuned tank to assume a positive potential, diode CR_1 will be forward biased. Conduction of the diode will cause current to flow in the direction indicated (arrow in Figure 36-1). The current will cause voltage drops across R_L and CR_1 with the polarities shown. The forward biased resistance of CR_1 will be a small percentage of the resistance of R_L. Thus, during the conducting half cycle of the circuit, the greatest percentage of the source voltage will appear across R_L as output voltage E_{out}. Due to the heterodyning action of the applied signals within the diode the output will be a composite waveform containing voltage pulses at many frequencies, of which the predominant components will be the original carrier and sideband frequencies plus their sums and differences. All of these voltage pulses are represented in the output waveform (Figure 36-1) as the resultant RF voltage pulses. Positive pulses only appear in the output waveform because CR_1 does not conduct when the resultant RF variations of the input signal are going through their negative alternations. The average output

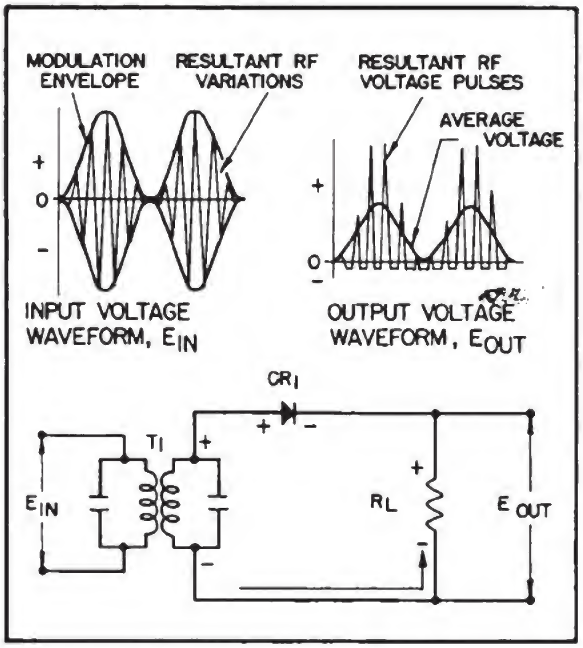

Figure 36-1 - Series diode detector with resistive load.

voltage (average of the resultant RF voltage pulses) will vary in accordance with the original modulating signal. Although there is a usable audio voltage component present in the output it is very small in magnitude.

Detector circuit efficiency (not to be confused with power efficiency) is determined by the amount of dc voltage developed across the load resistor for a given peak value of ac voltage input. This is expressed in the form of an equation as:

$$\%\text{ Rect. eff.} = \frac{\text{dc output volt.}}{\text{ac input volt.}} \times 100 \qquad (36\text{-}1)$$

Where:

% Rect. eff. = the percentage of circuit rectification efficiency

dc output volt. = the dc voltage developed across the diode load resistor.

ac input volt. = the peak amplitude of the applied ac voltage.

Equation (36-1) can be used to determine the efficiency of the circuit in Figure 36-1. For simplicity it will be assumed that the input to the detector is an UNMODULATED (constant amplitude) carrier of 10 volts peak amplitude. Since the forward resistance of CR_1 is very small in comparison to the resistance of R_L, virtually all of the rectified voltage will appear across R_L. For example, for an input of 10 volts peak the peak amplitude of the voltage pulses across R_L will be approximately 9.9 volts. The action of the diode detector circuit is similar to that of a half-wave rectifier power supply. Therefore, the average, or dc, output voltage will be:

$$E_{AV} = 0.318 \times E_{PEAK}$$

$$E_{AV} = 0.318 \times 9.9$$

$$E_{AV} = 3.15 \text{ volts}$$

Inserting values in equation (36-1).

$$\%\text{ Rect. eff.} = \frac{\text{dc output volt.}}{\text{ac input volt.}} \times 100 \qquad (36\text{-}1)$$

$$\%\text{ Rect. eff.} = \frac{3.15}{10} \times 100$$

$$\%\text{ Rect. eff.} = 31.5\%$$

It can be seen that the rectification efficiency of the diode detector with only a resistive load is very poor.

The efficiency of the circuit can be improved considerably by connecting a capacitor, of the proper value, in parallel with the load resistor. An arrangement such as this is shown in Figure 36-2. Before examining the circuit efficiency with the capacitor added a brief description of circuit action will be given.

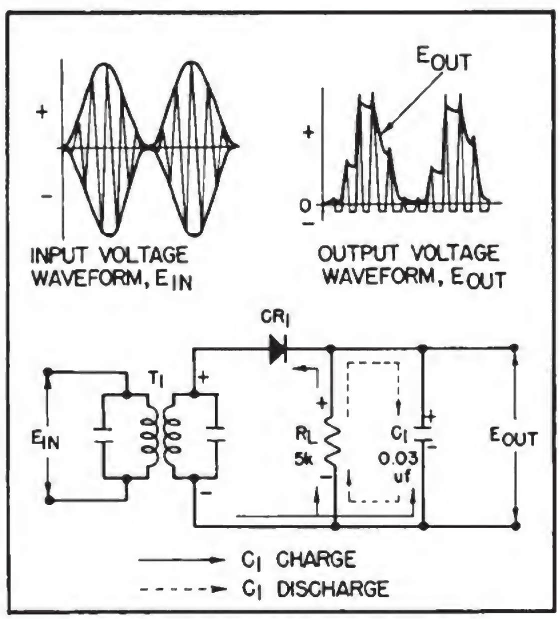

Figure 36-2 - Series diode detector with capacitor added.

When the input signal causes the top of T_1's secondary tank to become positive CR_1 will be forward biased. Current, indicated by the solid arrows in Figure 36-2, will flow from the bottom of the tank, through the parallel combination of R_L and C_1, through CR_1 and back to the top of the tank. Since the only resistance in the charge path of C_1 is the low resistance of the diode, C_1 will charge rapidly to almost the peak value of the voltage drop across R_L. As the input signal decreases and reverses polarity the voltage developed across the tank decreases, CR_1 will cease to conduct and C_1 will begin to discharge in an attempt to maintain the value of voltage established across R_L. However, the only path available for the discharge current of C_1 is through the load resistor R_L. The charge time of C_1 is short due to the low resistance of CR_1, while the discharge time is long due to the relatively large resistance of R_L. The pulses that control the conduction of CR_1 occur at the intermediate frequency. Therefore, C_1 will lose only a small percentage of its charge between pulses. This will cause the output voltage to follow the peak value of the rectified voltage very closely, as seen by the output waveform in Figure 36-2. It must be realized that the pulses in the waveform are expanded tremendously for the purposes of illustration. The sawtooth appearance, which depicts the charge and discharge of the capacitor, would not be visable in the actual output waveform.

Comparison of the output waveforms for a diode detector with, and without the capacitor shows that the voltage in both cases varies in accordance with the original modulating signal. However, the addition of the capacitor causes a greater magnitude of voltage and a more faithful reproduction of the modulating signal.

An example problem will be used to show how the addition of a capacitor improves the efficiency of the diode detector circuit. For simplicity it will again be assumed that the input to the detector is an UNMODULATED carrier of 10 volts peak amplitude. The voltage divider action of R_L and CR_1 will again cause the peak amplitude of the voltage across R_L to be approximately 9.9 volts. C_1 will charge very close to this value. Due to the action of C_1 and R_L in maintaining the value of voltage between charging pulses the average dc output voltage will remain at approximately 7.8 volts.

Inserting values in equation (36-1) yields:

$$\% \text{ Rect. eff.} = \frac{\text{dc output volt.}}{\text{ac input volt.}} \times 100 \qquad (36\text{-}1)$$

$$\% \text{ Rect. eff.} = \frac{7.8}{10} \times 100$$

$$\% \text{ Rect. eff.} = 78\%$$

774-639 O - 65 - 9

The rectification efficiency of the diode detector has been increased from 31.5% without C_1 to 78% with C_1. C_1 also acts as a filter to remove the high frequency components from the output.

Q1. Would the theory of operation for the series detector circuit remain the same if the diode connections were reversed?

Q2. Would increasing the ratio of R_1 to CR_1 improve the rectification efficiency?

Q3. What is the prime reason for the inclusion of C_1 (sometimes called the filter capacitor) in the detector circuit?

36-3. Choice of R and C Values

In all electronic circuits the choice of part values in usually a compromise between values for an ideal situation and a practical situation. Such is the case in choosing the values for R and C to be used in the output of the detector circuit. The rectification efficiency is highest when the value of R_1 is as large as possible in relation to CR_1. Filtering is best when the value of C_1 is large in comparison to the value of the diode capacitance. It would appear that the value of both R and C should be large. On the other hand one of the prime qualifications for the detector circuit, for faithful reproduction, is that the charge and discharge of C_1 be able to follow the modulation envelope as closely as possible. Thus, the discharge time constant of the circuit should be such that it appears long to the intermediate frequency (for good filtering and efficiency) and short to the modulating frequency (for faithful reproduction). Typical values of components used, in transistor broadcast band superhet receiver, to achieve a time constant exhibiting this quality are: C = 0.03 microfarads, R = 5 k ohms, and CR_1 = 50 ohms (actually the diode may range in value from 10 to a 100 ohms or more. Thus, 50 ohms is a representative value). The relationship between the charge time constant, discharge time constant, intermediate frequencies, and the modulating frequencies can be shown by using the typical component values in an example problem:

The detector circuit used in the example problem will consist of a 5k ohm load resistor, a 0.03 microfarad filter capacitor, and a diode with 50 ohms of dynamic resistance (Figure 36-2). The input signal is a composite waveform containing a 455 kc IF carrier and sideband frequencies of 454.6 kc and 455.4 kc (representing the original signal of 400 cps). For the circuit to operate properly the time constant for capacitor discharge should appear relatively long at the carrier frequency of 455 kc and

A1. Yes. The only difference being that the positive half cycles rather than the negative would be rectified.

A2. Yes. The larger R_1 is in comparison to CR_1 the larger the percentage of rectified voltage that will appear as E_{out}.

A3. Increase the efficiency of the circuit.

---

relative short at the modulating frequency of 400 cps. The charge time constant of the circuit must be very short. The charge time constant of the circuit is determined by the product of the diode resistance and the value of capacitance. Insert values:

$$T = RC \qquad (10\text{-}14)$$

$$T = 50 \times 3 \times 10^{-8}$$

$$T = 1.5 \times 10^{-6}$$

Thus, the time constant of the circuit during charge is 1.5 microseconds. The discharge time constant of the circuit is determined by the load resistance R_L and the value of capacitance. Insert values:

$$T = RC \qquad (10\text{-}14)$$

$$T = 5 \times 10^3 \times 3 \times 10^{-8}$$

$$T = 150 \times 10^{-6}$$

Thus, the time constant of the circuit during discharge is 150 microseconds.

In order to examine the relationship between the discharge time constant and the IF and modulating frequency it is necessary to determine the time for one cycle of each of these frequencies. The period for one cycle is equal to the reciprocal of the frequency. Therefore, the time for one cycle of the IF will be:

$$\text{Period} = \frac{1}{\text{frequency}} \qquad (8\text{-}5)$$

$$\text{Period} = \frac{1}{455 \times 10^3}$$

$$\text{Period} = 2.19 \text{ microseconds}$$

The time for one cycle of the modulating frequency will be:

$$\text{Period} = \frac{1}{\text{frequency}} \qquad (8\text{-}5)$$

$$\text{Period} = \frac{1}{400}$$

$$\text{Period} = 2500 \text{ microseconds}$$

From the results of the example problem it can be seen that the discharge time of 150 microseconds is long in respect to the 2.19 microseconds of the IF, but, is short in respect to the 2500 microseconds of the modulating frequency.

The correct choice of values for R and C in the diode detector circuit is important if maximum sensitivity and fidelity are to be obtained. If either R or C are made too large the time constant will be too long and the discharge of the capacitor will not follow the modulation envelope at the higher modulating frequencies. This will result in a situation such as illustrated in Figure 36-3.

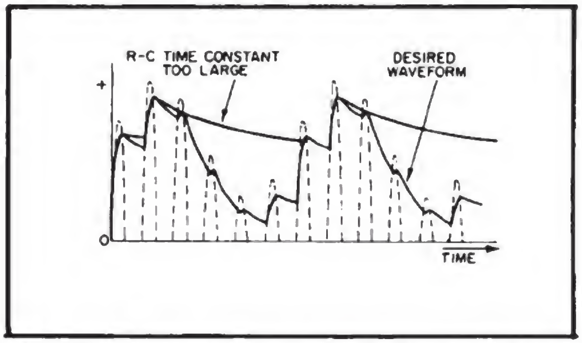

Figure 36-3 - Effect of too large a time constant.

From Figure 36-3 it can be seen that too long a discharge time constant will cause the negative peaks of the desired waveform to be clipped. This type of distortion is sometimes referred to as NEGATIVE PEAK CLIPPING or DIAGONAL CLIPPING.

Distortion will also be introduced if the discharge time constant is too short. Figure 36-4

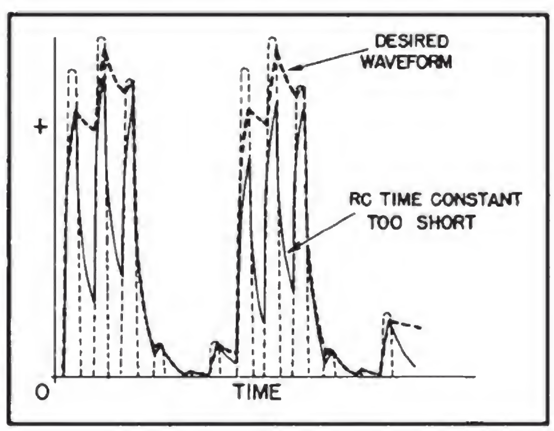

Figure 36-4 - Waveforms showing effect of too short a time constant.

illustrates the effect of too short a time constant. The dotted line in the figure indicates the desired waveform. The solid line (actual waveform) shows that the short time constant allows the capacitor to discharge considerably between pulses. The result is loss of rectification efficiency and the presence of a high frequency component in the output.

Q4. If the modulating frequencies were increased from the audio range to a much higher range, such as used in television or radar, would the discharge time constant of the detector circuit have to be decreased or increased?

36-4. Shunt Diode Detector

The SHUNT DETECTOR also produces a voltage and current output which is proportional to the modulation signal. The shunt detector is characterized by the diode being in shunt (parallel) with the load resistor rather than in series with it. Figure 36-5 illustrates a simple shunt detector with only a resistive load.

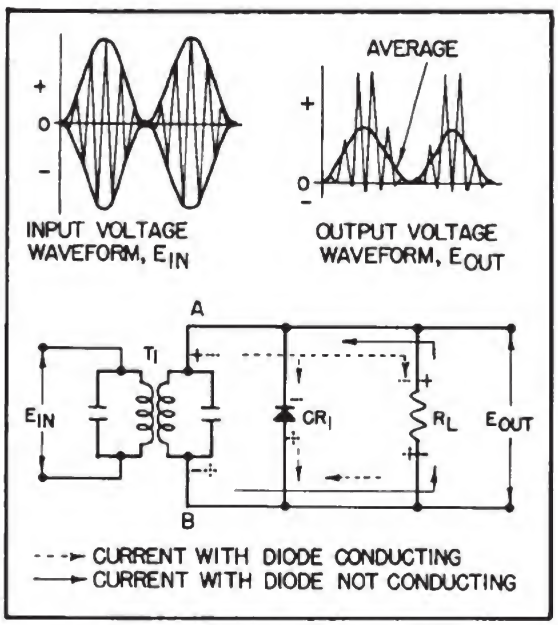

Figure 36-5 - Shunt detector with resistive load.

When the resultant RF variations of the input signal E_{in} cause point A to become positive with respect to point B (solid polarity signs) diode CR_1 will be reverse biased and will appear as a very high resistance in parallel with the load resistance R_L. Thus, the major current flow in the circuit will follow the path depicted by the solid arrow (Figure 36-5). This current flow will cause a voltage drop across R_L with the polarity (solid signs) shown. There will be a very slight current flow through the parallel path of CR_1 due to minority carrier flow, however, for all practical purposes its magnitude is so small that it is neglected. Therefore, the peak of the resultant voltage pulse developed across R_L is almost equal to the peak of the input voltage.

When the resultant RF variation of the input signal cause point A to become negative with respect to point B (dotted polarity signs) diode CR_1 will be forward biased and will appear as a very low resistance in parallel with the load resistor. As a result the path of major current flow (dotted arrow) will be through the diode. A very small percentage of the current will flow through the parallel path of R_L causing a very small voltage drop with the polarity shown by the dotted signs. This small reverse voltage drop is evident in the output waveform (Figure 36-5) by the extension of the voltage pulses slightly below the zero reference line. This action may be summed up by saying that when the diode is reversed biased the parallel combination of the CR_1 and R_L appear as a high resistance to the source, with a correspondingly high voltage drop, and when the diode is forward biased the parallel combination of CR_1 and R_L appear as a low resistance to the source, with a correspondingly low voltage drop. The end result is a rectification or distortion of the input signal which causes an average voltage (corresponding to the modulation signal) to appear in the output.

It can be seen that, as occured with the series detector and only a resistive load, the rectification efficiency of this circuit is also very poor.

The rectification efficiency of the shunt detector may be improved by the addition of an inductor in series with the load resistor. Figure 36-6 illustrates a shunt detector with an inductor added.

When point A becomes positive the current path is shown by the solid arrows and the voltage drops by the solid polarity signs. During this period of time the diode is in a non-conducting state and the major current flow is building a magnetic field around L_1. When point A becomes negative the source current (dashed arrows) flows through the diode and back to point B. By virtue of the collapsing magnetic field attempting to maintain current flow, L_1 now appears as a source for current flow through R_L. The path for current flow caused by the collapsing magnetic field of L_1 is shown by the dotted arrows (Figure 36-6). Notice that this current maintains a voltage drop of the same polarity as the original across R_L.

The values of L_1 and R_L are chosen with the same regard to the time constant as was shown in the series detector. In other words, the magnetic field of L_1 must not be allowed to collapse

A4. Decreased, in order for the discharge of the capacitor to follow the higher modulating frequencies.

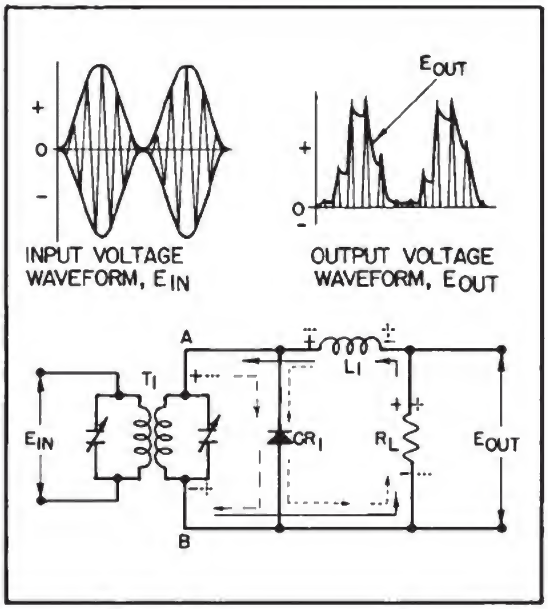

Figure 36-6 - Shunt detector with inductor added.

too much between RF current pulses and on the other hand it must collapse enough to follow the modulation envelope.

It must be remembered that the equation for the time constant of an RL circuit is slightly different than for an RC circuit. In an RL circuit the time constant is a direct function of L but an inverse function of R. This is shown by the equation:

$$T = \frac{L}{R} \qquad (9\text{-}13)$$

The effect of too long or too short a time constant on the output waveform of a shunt detector will be the same as in the series detector.

Q5. If the output were taken across the coil in a shunt detector, instead of across the load resistor, what frequencies would appear in the output? Explain.

Q6. How would circuit operation be effected if the coil were to become shorted?

36-5. Transistor Detector

It has been shown that the output voltage for a diode detector is always slightly less than the input voltage (rectification efficiency will never equal 100%). This slight loss of signal, due to detection, is of no concern in the average broadcast receiver because of the relatively large signal voltages present at the output of IF stage. However, there are many cases in electronics where it is necessary to detect a signal of very small amplitude. Two problems arise in the detection of small signals. First, since the signal is very small to begin with the slight loss due to detection is not desirable, and in some cases is intolerable. Second, it is necessary to use a device which is much more sensitive to small signal variations than the diode. In transistorized equipment this problem is solved by the use of a TRANSISTOR DETECTOR CIRCUIT. The transistor detector circuit combines the actions of detection and amplification within the same device.

A transistor detector connected in the common-emitter configuration is illustrated in Figure 36-7. Before discussing circuit operation a brief explanation of component functions will be given.

R_1 and R_2 constitute a voltage divider network to establish the no signal forward bias operating point. C_1, C_2, and C_3 are RF bypass capacitors and act to prevent the development of an RF voltage across the power supply, bias

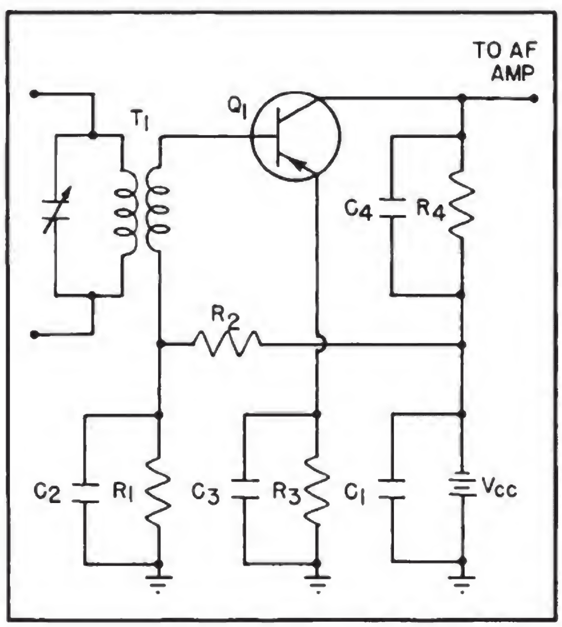

Figure 36-7 - Transistor detector.

resistor R_1, and emitter swamper R_3 respectively. Transformer T_1 couples the signal from the IF stage to the detector. Load resistor R_4 and demodulator filter capacitor C_4 perform the

function of audio signal development and RF filtering in the same manner as the load resistor and filter capacitor in the series diode detector. Transistor Q_1 is the device which performs the actions of detection and amplification. Application of a modulated IF signal to the primary of T_1 will cause voltage variations at the IF rate to be developed across T_1's secondary. The positive alternations of this induced signal will oppose forward bias and reduce collector current. The negative alternations will aid forward bias and increase collector current. Thus, the collector current will be caused to vary at an IF rate. Due to the heterodyning action of the various radio frequencies (present in the input signal) within the non-linear transistor the collector current variations will also contain a modulation component. As the collector current variations are applied to the parallel network R_4C_4 an action similar to that of the load resistor and filter capacitor network of the series diode detector takes place. The high frequency components are filtered out and a dc voltage, which varies in accordance with the original modulating signal, is developed between collector and ground.

The dynamic operation of the transistor detector circuit can best be demonstrated by use of the transfer characteristic curve. A typical curve (graph of collector current I_C versus voltage base-emitter V_{BE}) with waveforms is shown in Figure 36-8A.

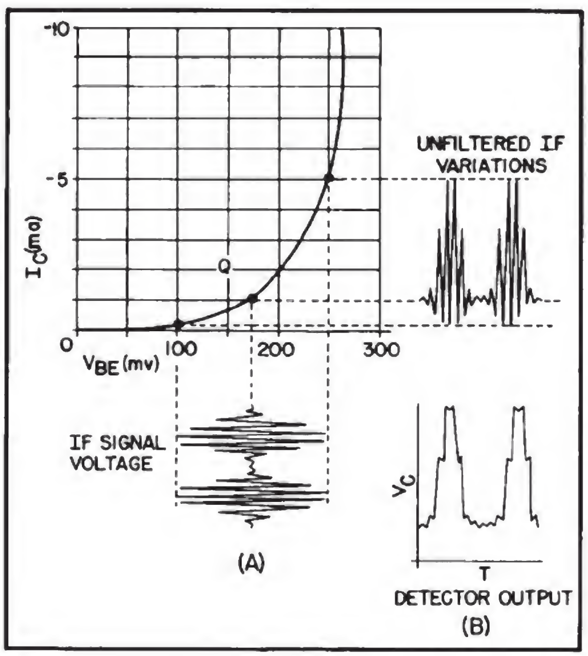

Figure 36-8 - I_C - V_{BE} curve with waveforms.

The graph depicts the extreme non-linearity of collector current with respect to base-emitter voltage. The symmetrical IF signal with 75 mv peaks is used as the input signal to the detector circuit. The Q point of the circuit is established with a base-emitter voltage of 175 mv and a collector current of 1 ma. Figure 36-8A shows that negative excursions (increase of V_{BE}) of the input signal produce much larger variations in I_C than do positive excursions (decrease of V_{BE}). Notice that the output waveform in Figure 36-8A is the collector current variations BEFORE application to the R_4C_4 network. The long time constant, at IF frequencies, and short time constant, at audio frequencies, of the network will convert the output to a voltage which very closely follows the peaks of the I_C variations. The output developed across the R_4C_4 network is shown in Figure 36-8B.

The main advantage of the transistor detector, over the diode detector, is its ability to provide a usable output signal with an extremely small input signal. The main disadvantage is that the transistor detector can not handle as large a signal input as the diode detector.

Q7. Could proper detection action be achieved if the transistor detector were biased on the linear portion of its characteristic curve? Explain.

Q8. What would be the result if too large an input signal were applied to the transistor detector?

36-6. Output Coupling

There are a number of methods of coupling the audio signal from the detector stage to the audio stage, such as RC coupling, transformer coupling, direct coupling, and impedance coupling. Each type of coupling has certain inherent advantages and disadvantages. Thus, the choice of type of coupling to be used for a particular application depends on the specific requirements of the circuit. In other words, some circuits may require a good frequency response while voltage gain is not particularly important, in other circuits high voltage gain may be the prime consideration while exceptional frequency response is not too important, in yet another circuit a compromise between these factors may be required.

RC and transformer coupling are the two basic methods and are the only ones to be considered in this chapter.

36-7. RC Coupling

The frequency response and voltage gain characteristics of RC coupling networks have been covered in great detail in section 19-34 and in

A5. Only radio frequencies would appear in the output since the coil offers a high impedance to RF but is almost a short to AF.

A6. The circuit would operate the same as if only the resistive load were present (loss of rectification efficiency, lower output voltage, high frequencies present in the output).

A7. No. Because heterodyning action would not take place.

A8. The transistor might be driven into saturation and distortion would result.

---

sections 20-18 through 20- 23. Since the use of RC coupling with transistors rather than electron tubes does not greatly alter the network characteristics only a brief review will be given here. If a more detailed review is desired the reader is directed to the above sections.

Figure 36-9 is a partial schematic illustrating the components involved in the RC coupling of a signal from the detector stage to the audio stage. The coupling components are included within the dotted area of the Figure. C_1 and R_1 have been discussed previously in connection with the series diode detector. These components are analogous to the plate load resistor and the electron tube output capacitance and effect the characteristics of the coupling network in approximately the same manner. C_2 is the coupling capacitor and performs the same function, with the same characteristics as the coupling capacitor in electron tube circuits. The parallel combination of R_2 and r_i (input junction resistance of the transistor) are analogous to the grid resistor. It must be realized that in comparing the action of coupling networks for electron tubes and semiconductor devices, that while the characteristics and actions may be the same, the component values will differ greatly.

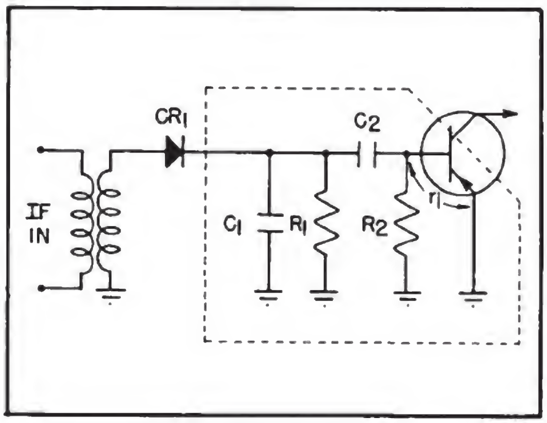

Figure 36-9 - RC coupling components.

Figure 36-10 illustrates a typical response curve of an RC coupling network. Depending on component values and design accuracy the low frequency limit may range from approximately 30 to 60 cps and the high frequency limit from approximately 100 kc to 200 kc.

The response is measured in terms of voltage gain over a range of frequencies. The gain falls off at very low frequencies because of the increase in the capacitive reactance of the coupling capacitor. This capacitor acts in series between the source and the load and has developed across it an increasing percentage of the signal voltage as the frequency is decreased. At midband frequencies the coupling capacitor is considered as practically a short. The reduction in gain at the high frequencies is caused by the shunting effect of the output capacitance, C_o, of one stage, the input capacitance, C_i, of the next stage, and the distributed capacitance of the coupling network. The combined effect of these capacitances is to increase the percentage of total signal voltage that is developed across the internal resistance of the input stage and decrease the percentage that appears as output voltage.

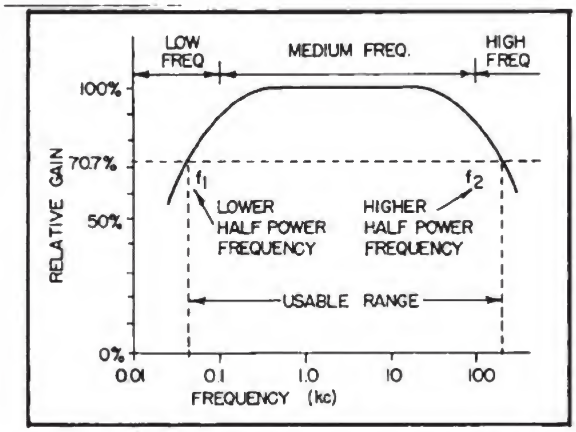

Figure 36-10 - Typical RC coupling frequency response curve.

36-8. Transformer Coupling

Diode detector circuits use RC coupling almost exclusively. On the other hand a transistor detector may use EITHER RC or transformer coupling. The common-emitter transistor detector (Figure 36-7) shown previously would, by necessity, use RC coupling, since connecting a low resistance transformer winding across the output would shunt the load resistance and V_{CC}. It is sometimes more desirable, in a transistor detector, to use transformer coupling rather than

RC coupling, due to the higher collector voltages obtainable.

Figure 36-11 illustrates a common-base transistor detector using transformer coupling. The non-linear and unilateral properties of the input junction provide the rectification (distortion) and heterodyning actions necessary for detection. Filtering of the high frequency component is provided by the C_2R_1 network, with additional high frequency filtering provided by C_3. The base-emitter voltage will vary in accordance with the audio modulation signal. This audio variation will be amplified and applied to the primary of T_2 and, thus, transformer coupled to the audio amplifier.

A transformer coupled stage has certain advantages over other types of coupling. The voltage amplification of the stage may exceed that of the transistor if the transformer has step-up turns ratio. Direct current isolation of the base, or emitter, of the next transistor is provided without the need for a blocking capacitor; and the dc voltage drop across the coupling resistor, which is necessary when RC coupling is used, is avoided. This type of coupling lends itself well to coupling a high impedance source to a low impedance load, or vice versa by choosing a suitable turns ratio.

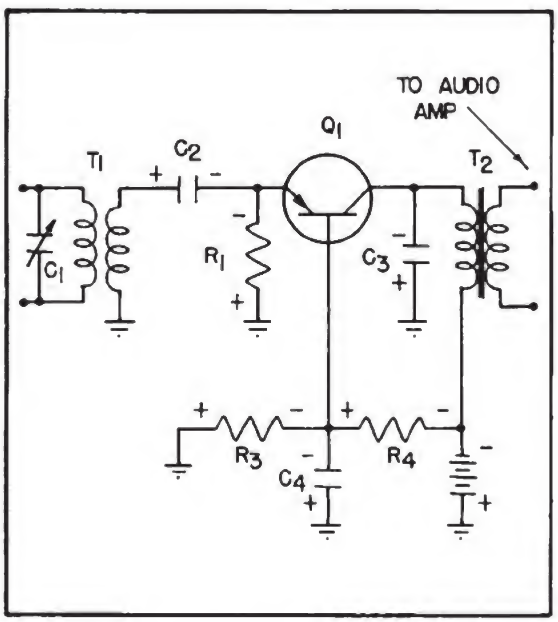

Figure 36-11 - Transformer coupled CB transistor detector.

Transformer coupling has the disadvantages of greater cost, greater space requirements, and the possibility of poorer frequency response at the higher and lower frequencies. The voltage gain as a function of frequency throughout the range in question is shown in Figure 36-12. The curve shows that the transformer coupled voltage amplifier has a relatively high gain and uniform frequency response over the middle range of audio frequencies, but poor response for both low and high audio frequencies.

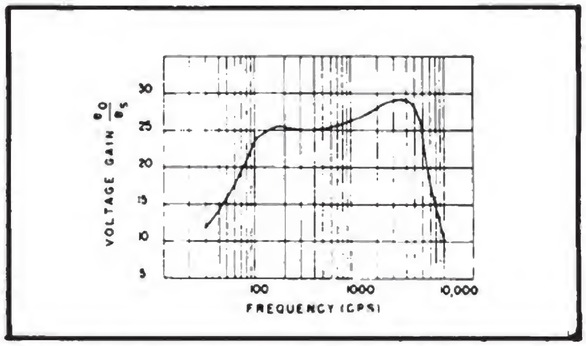

Figure 36-12 - Percents of voltage gain versus frequency-transformer coupling.

The decrease in response at the low frequencies is due to the decrease in inductive reactance of the transformer primary winding. The decrease in response at the high frequencies is due, mainly, to the shunting effect of transistor interelement capacitance (C_o and C_i), stray capacitances, and distributed capacitance of the transformer windings. If proper design precautions are not observed a peak in the response curve, such as observed in Figure 36-12 (around 4000 cps), will occur due to a series resonant condition being established between the leakage inductance and the stray capacitances. This series resonant condition usually occurs near the high frequency limit.

Q9. Why is the voltage gain of an RC coupled amplifier less than that of a transformer coupled amplifier?

Q10. What would occur (in respect to RF coupled to the audio amp) if a transformer coupled detector exhibited poor RF filtering?

AUTOMATIC VOLUME CONTROL

The function of an AUTOMATIC VOLUME CONTROL (AVC), also referred to as AUTOMATIC GAIN CONTROL (AGC), is to limit unwanted variations in the output of a receiver due to variations in strength of the received signal. To maintain a constant output level, a receiver without AVC would require readjustment of the manual volume control each time the strength of the received signal changes. Changes in signal strength occur as a result of changing stations and from fading caused by atmospheric conditions. Signals from stations operating at the same power may not reach the receiver antenna

A9. Because of I^2R losses and impedance mismatch.

A10. Due to the low gain of transformer coupling at high frequencies very little RF would be coupled to the audio amplifier.

---

with the same power, because of differences in transmission distances, carrier frequencies, atmospheric conditions, and obstructions between the transmitter and receiver antenna. The conclusion might be drawn that an AVC network is not necessary when the receiver is operating on one station, but the problem still exists, due to atmospheric conditions causing the signal strength to vary (fade in and out) or the antenna receives two signals which have traveled along different paths from the same transmitting antenna. For example, one signal may travel direct from the antenna, and the other signal may have been reflected from a distant object. Since the signal paths are constantly changing, the two signals will sometimes be in-phase and at other times be out-of-phase, thus tending to cancel or reinforce each other. The result is a variation in signal strength at the receiver antenna called fading. The effect of signal strength variations on the output power of an RF stage can best be demonstrated by an example problem.

Example. An RF amplifier, connected to a receiver antenna, has a power gain of 100. If the antenna receives an input signal of 10 micro-watts the output power (P_{out}) will be:

$$A_p = \frac{P_{out}}{P_{in}} \qquad (36\text{-}2)$$

Transposing:

$$P_{out} = A_p \times P_{in}$$

$$P_{out} = 100 \times 10 \times 10^{-6}$$

$$P_{out} = 1000 \text{ micro-watts}$$

The output power is equal to 1000 micro-watts and if fading is to be avoided the output power must remain at this level. However, if a reflected signal from the same station is received, of approximately half the strength (5 micro-watts) and in phase with the direct signal, the signal strength at the receiver antenna will increase to 15 micro-watts. To maintain the desired 1000 micro-watts of output power the power gain of the amplifier must be reduced. The new A_p will be:

$$A_p = \frac{P_{out}}{P_{in}}$$

$$A_p = \frac{1000 \times 10^{-6}}{15 \times 10^{-6}}$$

$$A_p = 67.7$$

When the 10 micro-watt direct signal and the 5 micro-watt reflected signal are out-of-phase, the signal strength at the receiver antenna will decrease to 5 micro-watts. To maintain the original 1000 micro-watt A_p, the power gain of the amplifier must be increased. The new power gain will be:

$$A_p = \frac{P_{out}}{P_{in}}$$

$$A_p = \frac{1000 \times 10^{-6}}{5 \times 10^{-6}}$$

$$A_p = 200$$

The variation of amplifier gain, in the example, to maintain a constant output power with a varying input signal can be accomplished automatically by the addition of an AVC circuit to the receiver. The purpose of the following sections is to show the methods and circuits used to produce AVC and the manner in which the AVC controls the receiver gain.

36-9. Simple AVC Circuit

The output of the detector circuit contains, among other components, a dc component which is proportional to the average carrier amplitude. The AVC circuit takes a portion of this dc component, filters it to remove the audio component, and applied it to the preceding stages. Figure 36-13 is a block diagram representation showing the AVC portion of the detector output being applied back to the preceding stages. The AVC

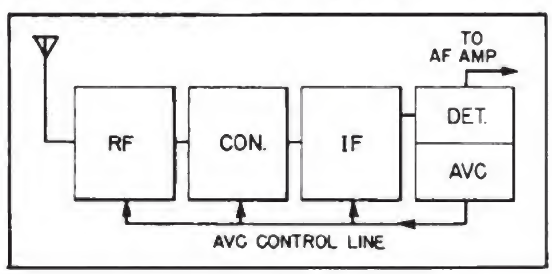

Figure 36-13 - Block diagram showing AVC application.

voltage may be used to control the amplification of all the stages preceding the detector stage. In an electron tube receiver the AVC voltage is usually negative, however, a transistor receiver may utilize either positive or negative voltage for AVC, depending on the type of transistors used and the element to which the control voltage is applied.

The schematic diagram of a simple AVC circuit used in conjunction with a series diode detector is shown in Figure 36-14.

The components T_1, CR_1, C_1, and R_1 constitute a normal series diode detector as previously described. The AVC network is composed of R_2 and C_2. In normal operation of the detector circuit with the potential shown, CR_1 conducts. Conduction of the diode will cause a charging current (shown by dotted line) to flow through the AVC capacitor (C_2) and AVC resistor (R_2). This charging current will cause C_2 to assume a polarity as shown. R_2 and C_2 form a voltage divider causing the voltage across C_2 to be only a portion of the voltage present across the output resistor R_1. When the potential across T_1 reverse biases the diode, and charging current ceases to flow, C_2 will begin to discharge. However, the discharge time constant of C_2, R_2, and R_1 is chosen to be longer than the lowest audio frequency present in the output of the detector. Consequently, C_2 will not discharge appreciably between peaks of the modulating signal and the voltage across C_2 will be a dc voltage. The voltage across C_2 is proportional to the average carrier signal. Thus, if the signal strength should vary, the average of the carrier signal will vary. This will cause C_2 to either increase or decrease its charge, depending on whether the signal strength increased or decreased. Since the charge of the AVC capacitor responds only to changes in the average signal level, instantaneous variations in the signal will not effect the AVC voltage.

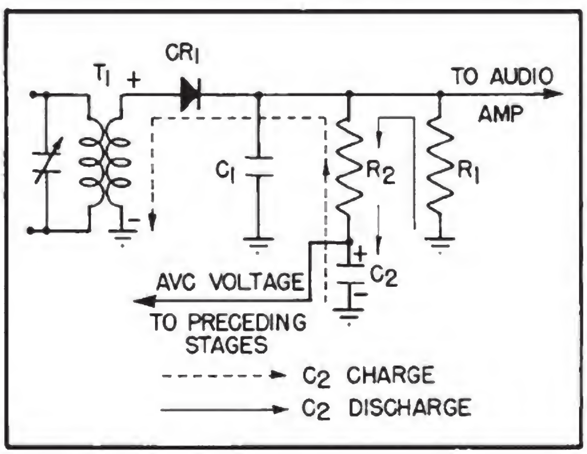

Figure 36-14 - Series diode detector and simple AVC circuit.

Q11. Why is a long time constant necessary in an AVC network?

36-10. Control of Amplifier Gain by AVC

It has been shown how a dc voltage that is proportional to, and will reflect the average variations of, the average signal level as obtained at the output of the AVC network. All that remains is to use this AVC voltage to control the amplification of one or more of the preceding amplifiers in the desired manner. Figure 36-15 illustrates a common-emitter amplifier with AGC (same as AVC) applied. The power gain of the common-emitter amplifier is shown, in this case, to be controlled by applying the AGC to the base element. A change in the AGC voltage will change the operating point of the transistor. This will change the dc emitter current and, hence, the power gain of the amplifier. In the circuit shown, R_1 and R_4 form a voltage divider network and establish the no-signal (forward) bias on the base. Since a PNP transistor is used the base has a negative potential. The AGC voltage from the detector is positive with respect to ground and is fed to the base through dropping resistor R_2. When the dc output of the detector increases (due to an increase in the average signal level) the AGC voltage will become more positive. This increased positive potential is applied to the base of Q_1, decreasing the forward bias of Q_1 and thereby decreasing the gain of the amplifier.

It may be noticed that AVC or AGC behaves, in this application, effectively as a controlled degenerative feedback. It should also be noted that use of an NPN transistor in the amplifier would have required that the AGC voltage posses a negative potential.

Q12. What could be considered as a disadvantage of the simple AVC circuit in respect to weak signals.

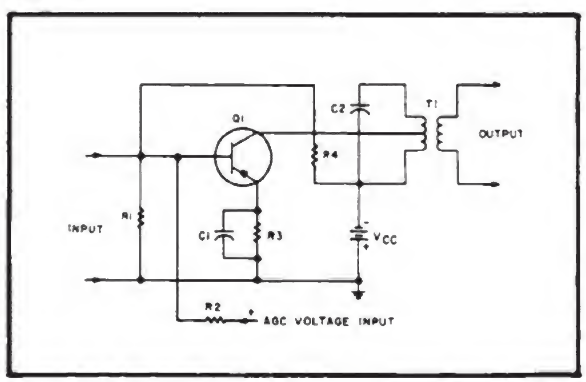

Figure 36-15 - CE amplifier with AGC.

A11. To prevent an audio signal component from being coupled back to the previous stages and producing undesired degeneration.

A12. Due to the degenerative action of the AVC voltage weak signals may receive insufficient amplification to provide a useful point.

EXERCISE 36

1. What purpose does a detector circuit serve in a superheterodyne receiver?
2. How is non-linear operation of a detector obtained and why is it necessary?
3. Is filtering required in detector circuits? What are the effects with and without filtering?
4. In what manner is the theory of operation of the detector circuit similar to the operational theory of the mixer stage?
5. Describe the operation of a series diode detector. Explain the result with each component opened and shorted.
6. What is meant by the % of rectification efficiency? How may it be improved in a given application?
7. What determines the output polarity of a detector circuit? How may the polarity be changed in a series detector? In a shunt detector?
8. Explain the manner in which the efficiency of the shunt diode detector is improved by the addition of the inductor.
9. Why is the sensitivity of the diode detector said to be low?
10. What are the basic requirements for the choice of RC values in a series diode detector? For RL in a shunt diode detector?
11. What will be the result if too large a resistance is used in the shunt detector?
12. Describe the purpose and action of the inductor in the shunt detector.
13. Describe two situations where the use of a transistor detector would be more advantageous than a diode detector.
14. Briefly explain the action of the filter capacitor and load resistor in the transistor detector.
15. Explain the characteristics of two basic methods of output coupling used in the detector stage.
16. Explain the action of an RC coupling network giving polarities and current paths for each component.
17. Explain the main function of AVC in a receiver. Give some reasons for needing AVC.
18. Briefly explain some of the factors necessitating the need for AVC.
19. What is the principle of operation of AVC circuits?
20. A series diode detector circuit uses a 0.4 megohm resistor for the diode load and a 200 picofarad capacitor for the filter. What is the time constant of the detector? Is this charge or discharge time?

CHAPTER 37

TRANSISTOR AF AMPLIFIER

Audio amplifier circuits are designed specifically to amplify signals which lie within the audio frequency spectrum (20 to 20,000 cps). In practice, to insure that these frequencies will have sufficient power, AF circuits are designed to amplify frequencies above and below the AF spectrum, often from 10 to 100,000 cps. Since transistors are essentially power-amplifying devices, their uses in audio circuits falls into two general categories: low level or high level audio amplifiers. The particular power level of an audio amplifier stage is determined by design requirements. In some cases (signal amplifiers) the power level ranges from picowatts to milliwatts, while for other cases (power amplifiers) the power level may be on the order of many watts.

This chapter will consider transistor audio amplifiers as used in typical superheterodyne receivers. The characteristics of low level amplifiers will be presented first, followed by a discussion of high level amplifiers. To simplify matters, the term "signal amplifier" will mean a low level power circuit, while the term "power amplifier" will designate a high level power circuit.

AUDIO SIGNAL AMPLIFIERS

To understand the relationship of the audio frequency section in a superheterodyne receiver to the other sections, refer to the block diagram of the superheterodyne receiver (Chapter 31). Note that the AF amplifier receives the demodulated output of the detector. The primary purpose for audio amplification is to provide sufficient power gain to drive the output transducer (speaker, headset, etc.). The power demands of the output transducer will determine the level of power reached in the AF section. If the AF amplifier output drives a headset or small speaker, the power required is small compared with that required to drive a combination of large speakers.

37-1. Single-Stage Audio Amplifier

The circuit of Figure 37-1 shows a basic singlestage audio amplifier.

Bias for the PNP, common-emitter, audio amplifier is established by the base-emitter current (solid arrows Figure 37-1). The current through the BE junction developes a voltage which forward biases the input circuit. Base resistor R_B limits the bias current thereby establishing the operating point of the transistor. Collector current is shown by the dotted arrows.

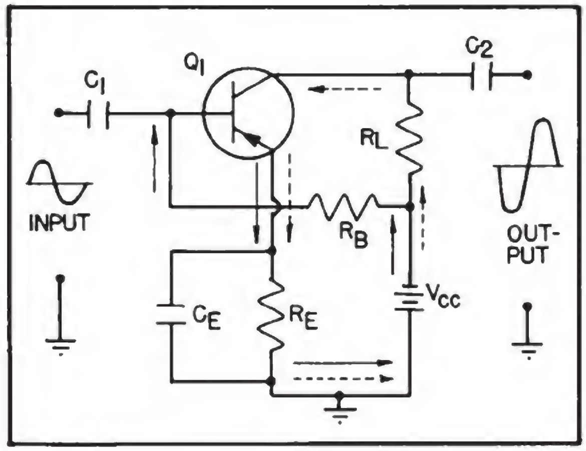

Figure 37-1 - Single-stage audio amplifier.

During the positive alternation of the input signal, forward bias is decreased. This causes a resultant decrease in collector current through R_L and collector voltage becomes more negative (with less of a voltage drop across R_L, the collector voltage will increase towards the negative value of V_{CC}).

During the negative alternation of the input signal the forward bias is increased. This increases the collector current through R_L, with the result that collector voltage becomes less negative.

C_1 is the input coupling capacitor. C_2 is the output coupling capacitor. R_E is the emitter stabilizing resistor and is paralleled by audio bypass capacitor C_E. The input and output voltage waveforms indicate a phase reversal and a voltage gain.

37-2. Effect of Input Resistance on Configuration choice.

When an audio amplifier is used to "build up" very weak audio signals it is often referred to

as a PREAMPLIFIER. Preamplifiers find their greatest application with weak audio signal sources such as microphones, phono pickups, tape heads, etc. The internal resistance of these sources may be either high or low depending on their type. For instance, phono pickups may be a crystal type having a high internal resistance or a dynamic type having a low internal resistance. For maximum transfer of signal the input impedance of the transistor used in the preamplifier should be close to the internal resistance (or impedance) of the device used as the source.

The most desirable method of matching source resistance to input resistance is by transformer coupling, however, this is not always practical. When the preamplifier must be fed from a low resistance source (20 to 1500 ohms), without benefit of transformer coupling, either the CB or CE configuration may be used. The CB configuration has an input impedance which is normally between 30 and 150 ohms; the CE configuration has an input impedance which is normally between 500 and 1500 ohms.

If the signal source has a high internal impedance (such as a crystal pickup head) then a high-input impedance preamplifier is also required. Assuming that transformer coupling cannot be used, it is possible to obtain the required high-input impedance by using one of the three following circuit arrangements.

It would appear that the easiest configuration to use would be the common-collector. The input resistance of the CC configuration is high because of the large negative voltage feedback in the base-emitter circuit. As the input voltage rises, the opposing voltage developed across the load resistance (R_L Figure 37-2) substantially reduces the net voltage across the BE junction. By this action, the current drawn

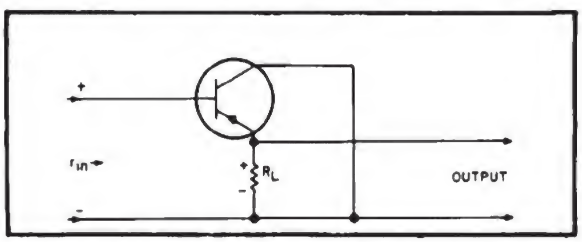

Figure 37-2 - Simplified schematic of CC preamp.

from the signal source remains low. By Ohm's law it is known that a low current drawn by a relatively high voltage represents a high resistance. If a load resistance (R_L) of 500 ohms is used, the input resistance of a typical CC configuration will be over 30,000 ohms. The disadvantage of the CC configuration, however, is that small variations in the current drawn by the following stage cause large changes in the input resistance value.

The variations of input resistance, as a function of load resistance, for the common-emitter, common-base, and common-collector configurations is shown in Figure 37-3. Notice that a common-collector configuration with an R_L of 500 ohms has an input resistance of approximately 35,000 ohms. Within the operating range of 1,000 to 100,000 ohms of load resistance the input resistance (r_i) of the CC circuit will increase with an R_L increase, r_i of the CE circuit will decrease with an R_L increase, and r_i of the CB circuit increases with an R_L increase.

The highest power gain is realized using a CE configuration. Its frequency response is adequate for use in most audio stages. The relationship of its input and output impedance are the most convenient of the three configurations, and when used with transformer coupling

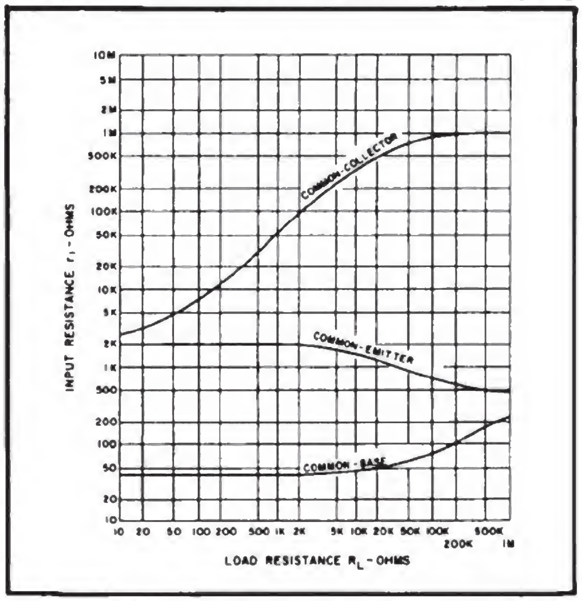

Figure 37-3 - Variations of r_i with R_L for each configuration.

permits easier impedance matching. Finally, its input resistance is less dependant on the value of load resistance than either the CB or CC configurations. Thus, the CE circuit is most often used for general purpose audio amplification.

The common-emitter configuration may be used to match a high source resistance by the addition of a series resistor in the base lead. The base-emitter junction resistance (represented by r_i in Figure 37-4) for a typical CE configuration is approximately 1,000 ohms, if a load resistance of 30,000 ohms is used. The input resistance, r_i, may be increased by reducing the load resistance (R_L). For instance,

decreasing the load resistance to approximately 10,000 ohms will increase r_i to approximately 1,500 ohms, as seen by the CE curve in Figure 37-3.

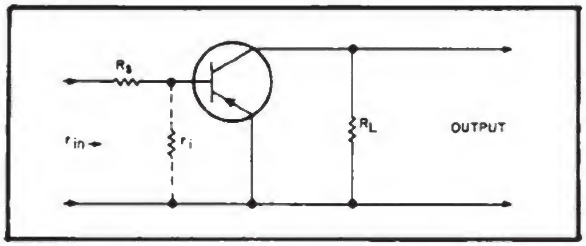

Figure 37-4 - Simplified schematic of CE preamp with series resistor.

Assume a source resistance of approximately 20,000 ohms. In order to use a CE configuration, having a 10,000 ohm R_L, it would be necessary to insert a series resistance (R_S Figure 37-4) of approximately 18,500 ohms in the base lead. Thus, the input resistance (r_{in}) of the amplifier will appear equal to the source. Notice, that before the series resistor was added, the input resistance r_i varied considerably with a change in R_L. When R_S is added to the circuit r_i will still vary with a change in R_L, however, r_i is now only a small portion of the total input resistance so that r_{in} remains essentially constant with load resistance changes. The fact that the total input resistance remains relatively constant even with large variations in transistor parameters or current drain by the following stage is the main advantage of this circuit arrangement. The disadvantage of this circuit, in addition to a small loss in current gain, is the large resistance in the base lead. This resistance leads to poor bias stability if the bias voltage is fed to the transistor through this resistor.

Another method of increasing the input resistance r_i of a common-emitter configuration is shown in Figure 37-5. This type of circuit is called a DEGENERATED CE configuration. If an unbypassed resistor (R_E Figure 37-5) is inserted in the emitter lead, the signal voltage developed across this resistor opposes the input signal voltage. As in the case of the common-collector configuration, this negative feedback voltage or degenerative voltage, causes an increase in the input resistance. With a bypassed resistor in the emitter lead the input resistance of the CE configuration would be 2,000 ohms, if a load resistor of 500 ohms were used (Figure 37-3). With an unbypassed resistor (R_E) of 500 ohms and a load resistor of 500 ohms, the input resistance (r_{in}) will appear as approximately 20,000 ohms. The input resistance may be made to appear as any desired value (within practical limits) by the proper choice of R_L and R_E. Like the CE circuit with the series resistor the total input resistance of the degenerated CE circuit will remain relatively constant with a varying load. However, the advantage of the degenerated CE configuration is that the unbypassed resistor (R_E) also acts as an emitter swamping resistor and helps to bias stabilize the transistor.

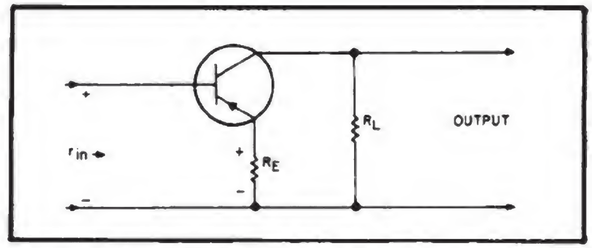

Figure 37-5 - Degenerated CE configuration.

Q1. What would be the approximate value of the series resistor inserted in the base lead of a CE amplifier, having an R_L of 50 k ohms, if it were desired to match its input resistance to that of a 5 k ohm source?

Q2. What advantage is gained by leaving the emitter swamper unbypassed, in a CE amplifier, when a high resistance source is used.

COUPLING METHODS

37-3. RC Coupling of Audio Amplifiers

As stated previously, a resistive-capacitive (RC) network is one of the common methods used to couple a signal from one stage to another. The RC network (shown within the dashed lines in Figure 37-6) used between two transistor stages consists of a collector load resistor (R_1) for the first stage, a dc blocking capacitor (C_1), and a dc return resistor (R_2) for the input element of the second stage.

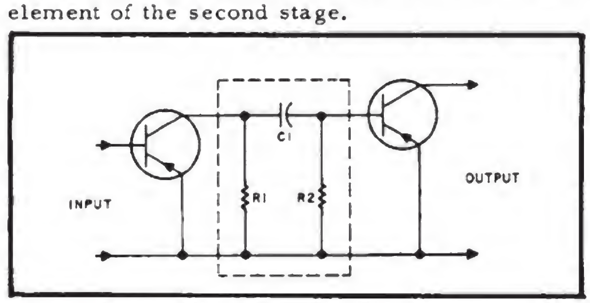

Figure 37-6 - RC Coupling network.

Because of the dissipation of dc power in the collector load resistor, the efficiency (ratio of ac power output to dc power delivered to stage) of the RC-coupled amplifier is low.

The dc blocking capacitor prevents the dc voltage of the collector of the first stage from

A1. R_S would be approximately 4200 ohms because with a 50 k ohm load r_i equals approximately 800 ohms (Figure 37-3).

A2. The degenerative voltage produced across the emitter swamper will increase the input resistance of the amplifier for impedance matching purposes.

appearing on the input terminal of the second stage. To prevent a large signal voltage drop across the dc blocking capacitor, the reactance of the capacitor must be small compared to the input resistance of the following stage with which it is in series. Since the reactance must be low then the capacitance value must be high. However, because of the low voltages used, the physical size of the capacitor can be kept small. The resistance of the dc return resistor is usually much larger than the input resistance of the second stage. This ratio is selected to prevent shunting the signal current around the input circuit of the second stage. The upper limit of the value of this resistor is dictated by dc bias temperature stabilization consideration (as discussed in Chapter 30).

The frequency response of the RC coupled audio amplifier is limited by the same factors that limited frequency response in other RC coupled circuits. In other words, the very low frequencies are attenuated by the coupling capacitor whose reactance increases with low frequencies. The high-frequency response of the transistor is limited by the shunting effect of the collector-emitter capacitance of the first stage and the base-emitter capacitance of the second stage. Circuit techniques that can be used to extend the low-frequency and the high-frequency response of RC-coupled transistors were discussed in Chapter 30.

RC coupling is used extensively in audio amplifiers, such as low level, low noise preamplifiers, due to the good frequency response, economy of circuit parts, and small size which can be achieved with it.

Q3. Why is a low value of capacitance undesirable in a coupling capacitor for an audio amplifier?

37-4. Transformer Coupling of Audio Amplifiers.

Interstage coupling of audio amplifiers by means of a transformer is shown in Figure 37-7. The primary winding of transformer T_1 (including the ac reflected load from the secondary winding) is the collector load impedance of the first stage. The secondary winding of transformer T_1 introduces the ac signal to the base and also acts as the base dc return path.

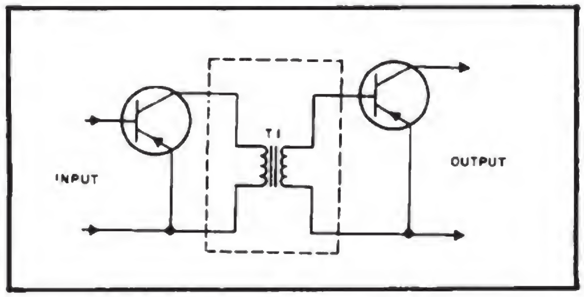

Figure 37-7 - Transformer coupling.

The very low resistance in the base path aids temperature stabilization of the dc operating point. With a swamping resistor in the emitter lead, the current stability factor is very good.

Because there is no collector load resistor to dissipate power, the power efficiency of the transformer-coupled amplifier approaches the theoretical maximum of 50 percent. For this reason, the transformer-coupled amplifier is used extensively in portable equipment where battery power is used.

Transformers facilitate the matching of the load to the output of the transistor and the source to the input of the transistor to bring about maximum available power gain in a given stage.

The frequency response of a transformer-coupled stage is not as good as that of the RC-coupled stage. The shunt reactance of the primary winding at low frequencies causes the low-frequency response to fall off. At high frequencies the response is reduced by the collector capacitance, and the leakage reactance between primary and secondary windings.

In addition, transformers are more expensive, heavier, and larger in size compared to resistors and capacitors required for coupling.

Q4. Why does the low resistance in the base path, offered by transformer coupling, aid temperature stabilization?

TRANSISTOR VOLUME CONTROL CIRCUITS

A necessary part of the audio frequency section is some means of manual control of the output signal level of the receiver. A MANUAL VOLUME CONTROL may be employed in a number of ways to control the volume level of a receiver. In a transistor receiver this control normally varies the amplitude of the signal applied to the input junction of the first audio stage.

37-5. Volume Control Requirements

NOISE. Volume controls should be so arranged in a given circuit that the volume control introduces no noise or a minimum amount of noise. This requirement can be achieved by avoiding the flow of dc current through the volume control.

GAIN. The volume control and its associated circuit should permit the variation of gain from zero to maximum.

FREQUENCY. The volume control should be so arranged in a given circuit that all frequencies are attenuated equally for all positions of the variable arm of the control.

37-6. Basic Volume Control Circuits

This section will discuss examples of several types of volume control circuits which do not meet the requirements listed in section 37-5.

The volume control circuit shown in Figure 37-8 DOES NOT meet the requirement of introducing a minimum amount of noise into the audio stages. In this application the variable resistor R_1 is used as the collector load for Q_1 and also the volume control. The dc collector current, of Q_1, flowing through the control will increase the noise produced by the circuit.

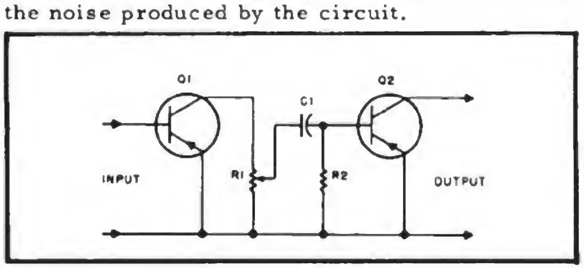

Figure 37-8 - Volume control-noisy.

The volume control circuit in Figure 37-9 eliminates the problem of dc current through the control, however, it is rarely used in receiver's because it does not meet the other requirements of a volume control.

1. The prime requirement of a volume control is to vary the gain from zero to maximum. Unless the variable resistor (R_2 in Figure 37-9) is made infinite in value, the amplifier gain cannot be reduced to zero.

2. The circuit increases the effective base response as more of variable resistor R_2 is used in the circuit. With the variable arm at the extreme left (zero resistance), capacitor C_1 used alone, will attenuate the lower frequencies more than the higher frequencies. As the variable arm moves to the right (increased resistance in circuit) the total coupling impedance will be determined mainly by the resistance. As a result, the

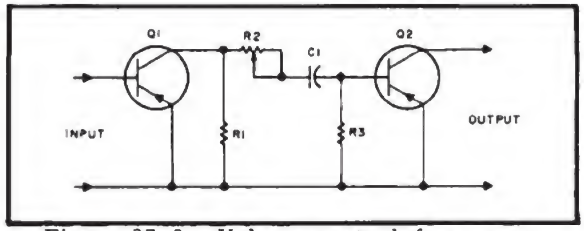

Figure 37-9 - Volume control-frequency discrimination.

relative attenuation of the high frequencies compared to the low frequencies will be greater.

Another method of controlling the volume of a transistor receiver is shown in Figure 37-10. This circuit is called a DEGENERATIVE EMITTER volume control. This method of volume control is rarely used, because it violates all three of the requirements in section 37-5, and is shown merely as an additional method which might find an application under certain conditions. The manner in which it fails to meet the requirements are listed below:

1. Unless variable resistor R_1 is made extremely large (20,000 to 50,000 ohms) the signal cannot be reduced to zero. A resistance of this value would require a very high value of battery V_{EE}) voltage.

2. The heavy emitter current flowing through variable resistor R_1 causes excessive noise.

3. The circuit increases the effective base response as more of variable resistor R_1 is bypassed by capacitor C_1. This action occurs because capacitor C_1 becomes more effective in bypassing the low frequencies as the resistance it must bypass increases. This results in less degeneration of the low frequencies than the highs.

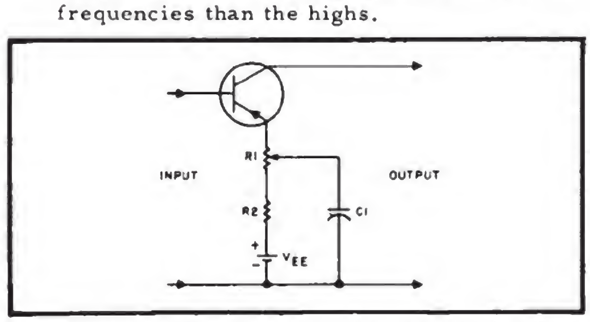

Figure 37-10 - Volume control-degenerative emitter control.

A practical method of volume control is shown in the two stage amplifier of Figure 37-11. The

A3. The high value of reactance at low frequencies causes an excessive loss of signal voltage to the following stage.

A4. If the base path contains a large resistance, even the very small current due to I_{CBO} can cause a significant voltage drop, thereby, changing the bias. Transformer coupling keeps this undesirable voltage drop to a minimum.

---

two stage amplifier uses variable resistor R_2 as the volume control. Collector load resistor R_1 developes the output signal of transistor Q_1. Capacitor C_1 couples the signal from Q_1 to the parallel combination of R_3, variable resistor R_2, and the input junction of Q_2. Resistors R_4 and the parallel combination R_2 and R_3 form a voltage divider to establish the potential applied to the base of Q_2. Resistor R_5 is the emitter swamping resistor and is ac bypassed by capacitor C_2. Collector load resistor R_6 developes the output signal of Q_2.

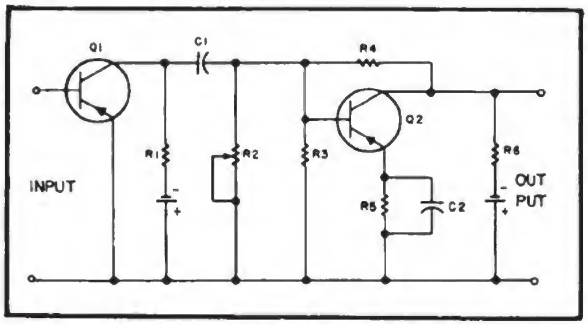

Figure 37-11 - Volume Control - two stage amplifier.

Although this circuit violates the requirement of no dc current through the volume control it should be noted that the path for this current is through the relatively large collector load R_6, the relatively large stabilizing resistor R_4, and the parallel combination of R_3 and the volume control R_2. Thus, the amount of dc current through R_2 will be comparatively small and will not increase the noise level of the amplifier by a significant amount.

Resistor R_2 controls the volume of the receiver by varying the base-emitter bias of Q_2. BE bias will vary from a maximum to zero depending on the position of the variable arm on resistor R_2. This condition can be understood by considering the base-emitter junction resistance of Q_2 as negligibly small. With this assumption, variable resistor R_2 can be considered as shunting the parallel combination of C_2, R_3, and the emitter swamper R_5. When the variable arm of R_2 is set to the maximum resistance (bottom end) position the maximum base voltage will be applied. When the arm of R_2 is set to the zero resistance (upper) end R_3 and C_2R_5 are effectively shunted by a zero resistance path. This places the base and emitter at the same potential, reducing the gain of Q_2 to practically zero.

Although this is not a particularly desirable method of controlling receiver volume it is very effective and, due to cost considerations, is extensively used in less expensive receivers.

Q5. What undesirable effect would the use of the degenerative emitter volume control circuit have on the input resistance of the amplifier?

AUDIO POWER AMPLIFIER

A situation that generally exists in the audio amplifying section of a receiver, or in similar sections of low frequency equipment, is to have a series of amplifiers that take a small signal and increase it until it has sufficient power to perform some useful work. Regardless of whether the audio section contains few, or many, amplifiers between the input and output the function of each stage is to increase the signal level until the final stage is reached.

It was stated previously that "signal amplifier" would indicate a low level power amplifier, while "power amplifier" would indicate a high level power amplifier. However, an additional distinction of the term "signal amplifier" must be made at this time. Signal amplifiers will be further qualified as SMALL-SIGNAL amplifiers and LARGE-SIGNAL amplifiers. The distinction is necessary because many times an audio section contains more than just a small-signal amplifier and a power-output stage. In these cases the stages following the small-signal amplifiers are not strictly power amplifiers, but, do work at a higher signal level than the first stages so that the term small-signal amplifier no longer applies. Hence, the term "large-signal" amplifiers to indicate a MEDIUM power level stage.

In general, the LAST stage of a series of audio amplifiers is called the POWER STAGE. The power stage differs from the preceeding stages in that it is usually designed to obtain MAXIMUM POWER OUTPUT rather than maximum power gain.

The discussion of audio power amplifiers will begin with phase inverter circuits, which provide the required input signals for proper operation of push-pull amplifiers. Then the single-ended power amplifier, and finally various push-pull power amplifier circuits will be considered.

37-7. Phase Splitters and Inverters

DRIVER STAGE is the term used to describe the amplifier which is used to supply the "driving", or input, signal to the final, or power amplifier stage. In the cases where the power amplifier contains a push-pull circuit it is necessary to use a PHASE SPLITTER or PHASE INVERTER as the driver stage. The purpose of a phase splitter, or inverter, is to supply two equal amplitude output signals, differing in phase by 180° with respect to each other, from a single input.

The various methods used to accomplish this purpose is the subject of this section.

Through the use of a center-tapped-secondary transformer, a single-ended audio stage can produce the signal requirements for push-pull operation. Figure 37-12 shows an audio driver stage with a transformer phase splitter.

In Figure 37-12 the audio signal is coupled from the volume control (R_1) to the input of the driver stage by C_1. Resistors R_2 and R_3 establish fixed base bias while R_4 and C_2 provice emitter bias. As the input signal goes through its positive alternation, the forward bias of Q_1 decreases with a resultant decrease in transistor conduction. Collector current, from V_{CC} through T_1 primary decreases and the collector potential (top of T_1 primary) becomes more negative. Through transformer action, the bottom of the secondary goes negative (with respect to the secondary center tap) while the top of the secondary goes positive.

When the input signal goes through its negative alternation the forward bias of Q_1 is increased. Transistor conduction increases, and the top of T_1 primary goes in a positive direction (less negative). Through transformer action the bottom of T_1 secondary produces a positive output whereas the top of the secondary produces a negative output. It can be seen that the secondary signals are of equal amplitude, but 180° out of phase, satisfying the input signal requirements of a push-pull power amplifier.

Although the transformer phase splitter provides a simple means of developing the required signals, economy, size, and weight may dictate the use of other means of developing the required signals.

Figure 37-13 shows a SPLIT-LOAD phase inverter which developes two signals, 180° out of phase, without the use of a transformer.

Transistor Q_1 output current flows through collector load resistor R_3 and emitter load resistor R_2. Resistors R_2 and R_3 are equal in value. Resistor R_1 establishes the base bias voltage. When the input signal aids the forward bias (base becomes more negative), the output current (I_o) increases. The increased output current causes the top side of resistor R_3 to become more positive with respect to ground, and the top side of resistor R_2 to become more negative with respect to ground. When the input signal opposes the forward bias, the output current decreases and causes voltage polarities across resistors R_3 and R_2 opposite to those indicated. This action produces two output signals that are reversed 180° with respect to each other.

In this circuit, equal voltage output are obtained by making resistor R_2 equal in value to resistor R_3. However, an unbalanced impedance results because the collector output impedance of transistor Q_1 is higher than its emitter output impedance. This disadvantage is overcome by the addition of a series resistor between C_2 and the top of R_2. The values of R_2 and the series resistor are chosen so that the output impedances of the collector and emitter are balanced. This eliminates distortion at strong signal currents. The signal voltage loss across the series resistor is compensated by making R_2 higher in value than R_3.

Notice that emitter resistor R_2 is unbypassed

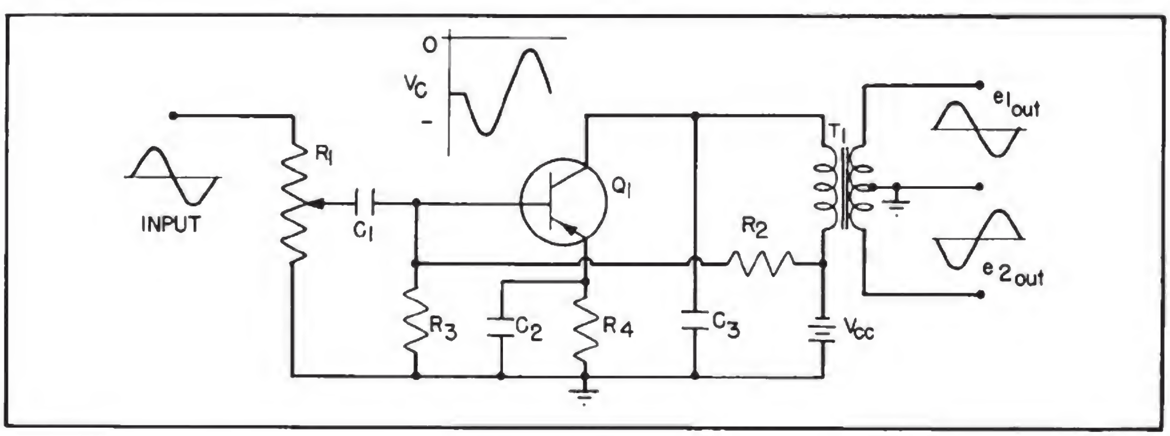

Figure 37-12 - Driver stage with transformer phase splitter.

A5. As the volume control is varied the amount of unbypassed resistance in the emitter circuit changes. This will cause the input resistance of the amplifier to vary (section 37-2).

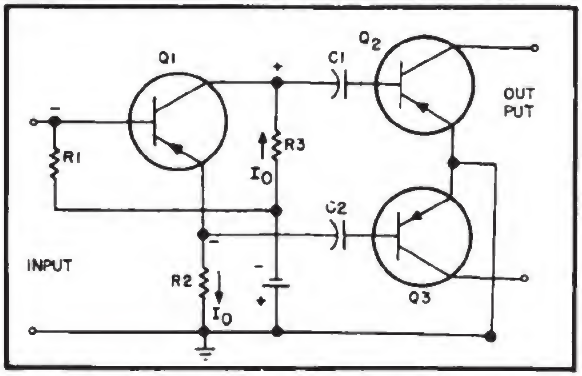

Figure 37-13 - Split-load phase inverter.

in order to develop one of the output signals. Because of the large negative feedback voltage developed across resistor R_2, a large signal input is required to drive the one-stage phase inverter. This disadvantage can be overcome by using two-stage phase inverters. In addition, a two-stage phase inverter provides more power output than a one-stage phase inverter. This advantage is important if the driver stage must feed a large amount of power to a high-level push-pull power output stage.

Figure 37-14 shows a two-stage phase inverter consisting of two identical CE configurations. Two signals, 180° reversed in phase, are obtained from the phase inverter. Assume that an input signal drives the base of transistor Q_1 negative. Because of the 180° phase reversal in the CE configuration, the transistor Q_1 collector goes positive. One portion of this positive signal is coupled to the base of transistor Q_2 through dc blocking capacitor C_2 and attenuating resistor R_4. The other portion of the positive signal is coupled through the dc blocking capacitor C_4 to one input circuit of a push-pull output stage. The positive-going signal on the base of transistor Q_2 causes a negative-going signal at the collector of transistor Q_2. This negative signal is coupled through dc blocking capacitor C_5 to the other input circuit of a push-pull output stage.

Resistor R_1 provides base bias for transistor Q_1. Collector load resistor R_3 develops transistor Q_1's output signal. Resistor R_2 is the emitter swamping resistor and is ac bypassed by capacitor C_1. Resistor R_5 provides base bias for transistor Q_2. Collector load resistor R_6 develops transistor Q_2 output signal. Resistor R_7 is the emitter swamping resistor and is ac bypassed by capacitor C_3.

Since two identical CE configurations are used, the source impedances are equal for the two input circuits of the push-pull output stage. In addition, the amount of power that can be delivered by the two-stage phase inverter is much greater than that of the split-load phase inverter.

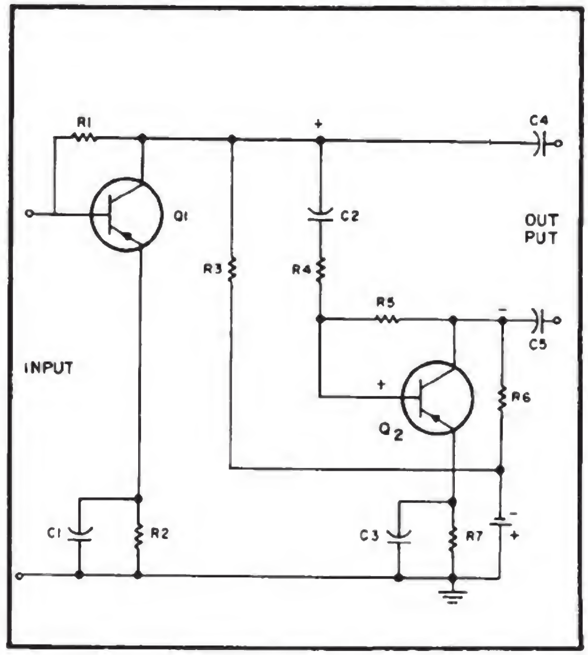

Figure 37-14 - Two stage phase inverter.

Q6. Why must a resistor be added in series with the emitter output coupling capacitor in a split load phase inverter?

Q7. If a push-pull output stage required a large input power, would a one-stage or two-stage phase inverter be used?

Q8. Could impedance matching of the inverter stage to the push-pull stage be accomplished easier by use of the transformer phase splitter or the split-load phase inverter?

37-8. Single-Ended Power Amplifier

The prime function of a power amplifier is to produce maximum output power. Due to transistor operation limitations (such as maximum current, voltage, and power dissipation ratings) the conditions for maximum power gain do not necessarily coincide with those for maximum power output. The maximum power dissipation (P_{Dmax}) rating of a transistor is very important

in the operation of a power amplifier, for it is this rating which limits the power output obtainable from any specific transistor.

For all practical purposes the schematic diagram of a power amplifier is similar to that of any low-power or medium-power amplifier. The major difference being the higher power rating, large physical construction, and mounting methods of the power transistor. One other major difference (aside from the fact that the power amplifier is the last stage) is that the power amplifier, being designed for maximum output power rather than maximum gain, will usually have a much smaller value of load impedance than do the preceding stages.

One requirement of any audio amplifier is minimum distortion. In order for a single-ended power amplifier to produce minimum distortion it must be operated class A. In order to obtain maximum output power, with minimum distortion, the load line of a class A power amplifier must lie tangent to the constant power-dissipation curve at a point where the mid-value of bias current crosses the curve. Meeting this requirement will permit a maximum swing of the input signal current.

Figure 37-15 illustrates a simplified schematic of a single-ended, class A, power amplifier. All components perform basically the same function as in any medium-power amplifier.

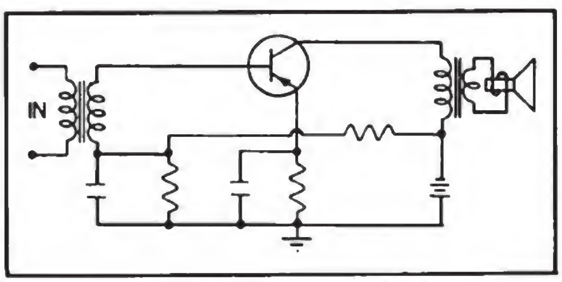

Figure 37-15 - Class A, single-ended, power amplifier.

Q9. Could a power transistor with a P_{Dmax} rating of 5 watts have a Q point at $V_C = 2$ V and $I_C = 666$ ma? Explain.

Q10. Would maximum output power be obtained from a class A power amplifier, with a P_{Dmax} rating of 3.5 watts, if the Q point were at $V_C = 6$ V and $I_C = 433$ ma? Explain.

37-9. Class A, Push-Pull, Power Amplifier

The class A push-pull amplifier consists essentially of two transistors connected back to back, with both transistors biased for class A operation. Figure 37-16 shows a simplified circuit for a class A push-pull power amplifier.

Through the use of a transformer phase splitter (T_1), two signals, 180° out of phase are applied as inputs to the push-pull amplifier. R_1 limits the base bias current to establish the desired operating point of Q_1 and Q_2. One half of the primary of transformer T_2 represents the collector load impedance for Q_1 while the other half represents the collector load for Q_2. T_2 also provides impedance matching between the relatively high output impedances of the transistors and the low impedance of the speaker voice coil.

On the positive alternation of the input signal the potential on the base of Q_1 will increase in a positive direction, while the potential on the base of Q_2 will increase in a negative direction. Since both transistors are of the PNP type, the potentials applied to their respective base elements will cause the conduction of Q_1 to decrease and the conduction of Q_2 to increase.

Current flows from the center tap of T_2 TOWARDS point B. Under quiescent operating conditions these currents are equal and the magnitude and polarity of the voltage drops they produce are such that there is no difference of potential between points A and B. The center tap of T_2 is fixed at the maximum negative value in the circuit by virtue of being connected to V_{CC}. Thus, the potentials at points A and B are less negative by some amount than the center tap. As stated, the positive alternation of the input signal causes the conduction of Q_1 to decrease. In other words, point A becomes more negative

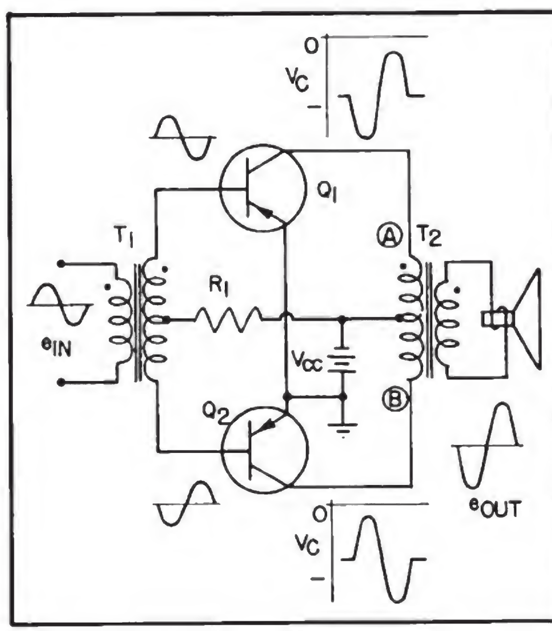

Figure 37-16 - Class A, push-pull, power amplifier.

A6. In order to make the emitter output impedance equal to the collector output impedance.

A7. Two stage.

A8. Transformer phase splitter.

A9. No. Operating point would fall beyond the constant power dissipation curve (8 watts).

A10. No. The load line would not be tangent to the P_{Dmax} curve. Power output would only be 2.6 watts.

---

because it is approaching the value of the potential at the center tap. The collector voltage waveform of Q_1, which is actually a graph of the potential at point A is seen to be increasing in a negative direction at this time.

The positive alternation of the input signal also causes the conduction of Q_2 to increase. This causes the potential difference between point B and the center tap to increase. Thus, the potential at point B is increasing in a positive direction (as shown by the collector voltage waveform of Q_2). Since point A is becoming more negative and point B is effectively becoming more positive, there is a potential difference developed across the entire primary winding. By transformer action this potential difference is coupled to the secondary of T_2 and appears as the waveform e_{out}.

On the negative alternation of the input signal the reverse of the above actions occurs, as can be seen by the collector voltage waveforms of Q_1 and Q_2, and the polarity of the voltage developed across the primary of T_2 is reversed.

The power output from this class A push-pull circuit is more than twice that obtainable from a single ended, class A, power amplifier. An added advantage of this circuit is that, due to the push-pull action of the output transformer, all even harmonics are eliminated in the output (if the transistor circuits are balanced). Since the distortion is caused mainly by second harmonics, elimination of these harmonics will result in a relatively distortion-free output signal.

The class A push-pull power amplifier finds its greatest application where minimum distortion is the primary consideration and high output power and efficiency are deemed less important.

Q11. Give two main advantages of the class A push-pull amplifier over the single stage power amplifier.

37-10. Class B, Push-Pull, Zero Bias Amplifier

Figure 37-17 shows a simplified circuit of a class B amplifier. The emitter-base junctions are zero biased. In this circuit each transistor

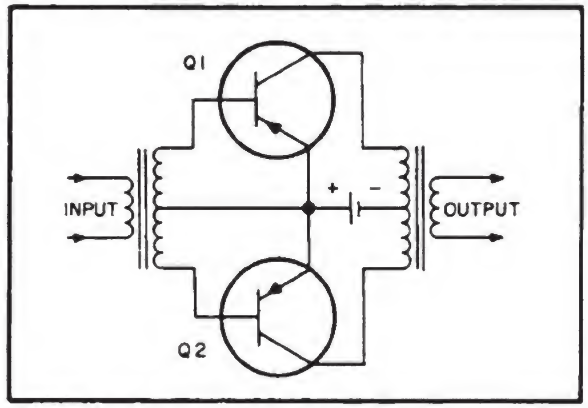

Figure 37-17 - Class B, push-pull amplifier with zero input bias.

conducts on alternate half cycles of the input signal. The output signal is combined in the secondary of the output transformer. Maximum efficiency is obtained even during idling (no input signal) periods, because neither transistor conducts during this period.

An indication of the output current waveform for a given signal current input can be obtained by considering the dynamic transfer characteristic for the amplifier. It is assumed that the two transistors have identical dynamic transfer characteristics. This characteristic, for one of the transistors, is shown in Figure 37-18A. The variation in output (collector) current is plotted against input (base) current under load conditions. Since two transistors are used, the overall dynamic transfer characteristic for the push-pull amplifier is obtained by placing two of the curves (Figure 37-18A) back-to-back. The two curves are shown back-to-back and combined in Figure 37-18B.

Note that the zero line of each curve is lined up vertically to reflect the zero bias current. In Figure 37-19, points on the input base current (a sine wave) are projected onto the dynamic transfer characteristic curve. The corresponding points are determined and projected as indicated to form the output collector current waveform. Note that severe distortion occurs at the crossover points; that is, at the points where the signal passes through zero value. This is called crossover distortion. This type of distortion becomes more severe with low signal input currents. Crossover distortion can be eliminated by using a small forward bias on both transistors of the push-pull amplifier.

37-11. Class B, Push-Pull, Low Bias Amplifier

A class B push-pull amplifier with a small forward bias applied to the base-emitter junctions is shown in Figure 37-20A. A voltage divider is formed by resistors R_2 and R_1. The voltage developed across R_1 supplies the base-emitter bais for both transistors. This small forward bias eliminates crossover distortion.

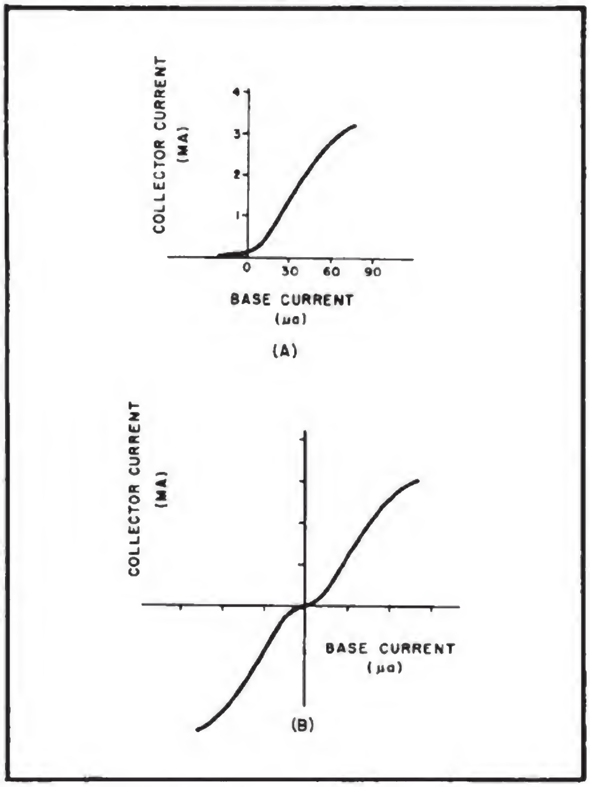

Figure 37-18 - Dynamic transfer characteristic curves.

In the voltage divider (A, Figure 37-20), electron current flow from the battery is in the direction of the arrow. This current establishes the indicated polarity across resistor R_1 to furnish the required small forward bias. Note that no bypass capacitor is used across resistor R_1. If a bypass capacitor were used (B, Figure 37-20), the capacitor would charge (solid-line arrow) through the base-emitter junction of the conducting transistor (during the presence of a signal, and discharge (dashed-line arrow) through resistor R_1. The discharge current through resistor R_1 would develop a dc voltage with the polarity indicated. This is a reverse bias polarity that could drive the amplifier into class C operation with the resultant distortion even more severe than crossover distortion. The capacitor must not be used.

A study of the dynamic transfer characteristic curve of the amplifier demonstrates the

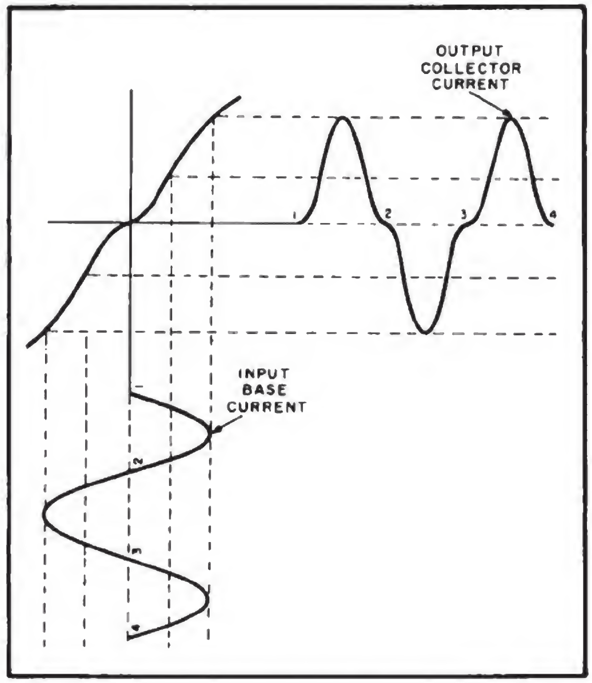

Figure 37-19 - Dynamic transfer characteristic curves of class B, push-pull amplifier with zero bias, showing input and output current waveforms.

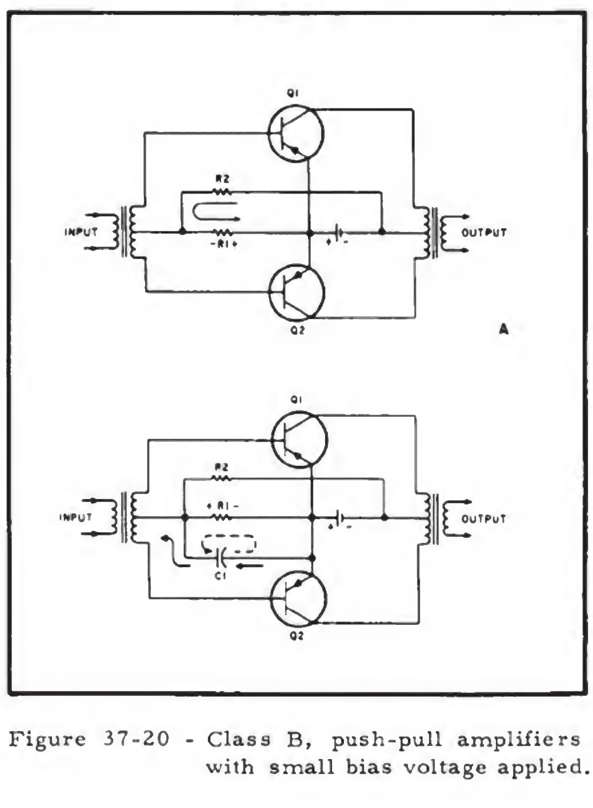

Figure 37-20 - Class B, push-pull amplifiers with small bias voltage applied.

A11. More output power and less distortion at high signal levels.

---

elimination of crossover distortion. In A, Figure 37-21, the dynamic transfer characteristic curve of each transistor is placed back-to-back for zero base current bias conditions. The two curves are back-to-back and not combined. The dashed lines indicate the base current values when forward bias is applied to obtain the overall dynamic characteristic curve of the amplifier. With forward bias applied, the separate

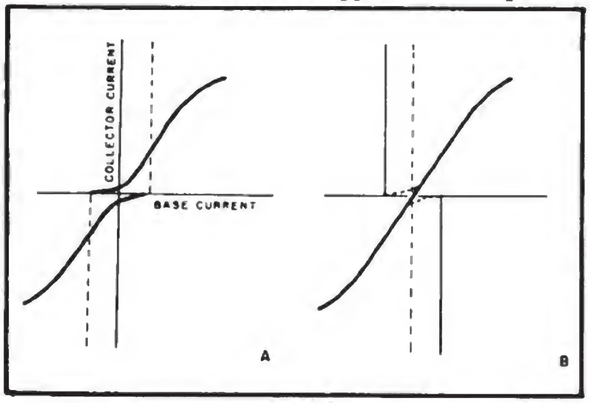

Figure 37-21 - Dynamic transfer curves, low bias, class B.

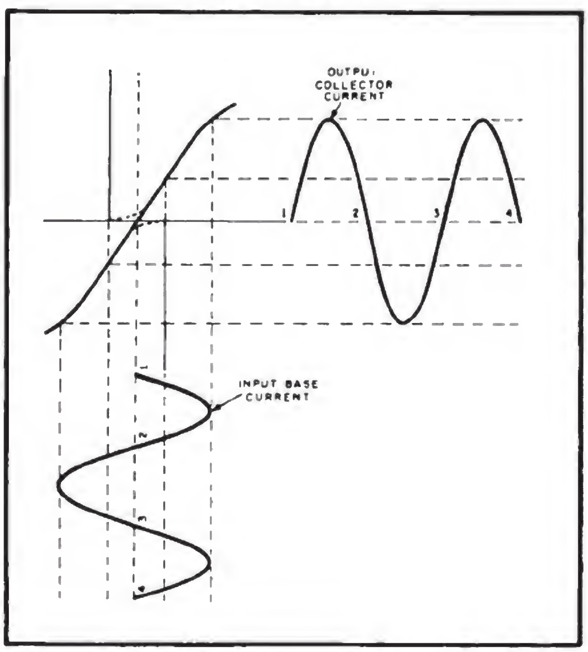

Figure 37-22 - Dynamic transfer characteristic curves of class B, push-pull amplifier with small forward bias, showing input and output current waveforms.

curve of each transistor must be placed back-to-back and aligned at the base bias current line (dashed line). The zero base current lines (solid lines) are offset (B, Figure 37-21).

In Figure 37-22, points on the input base current (a sine wave) are projected onto the dynamic transfer characteristic curve. The corresponding points are determined and projected as indicated to form the output collector current waveform. Compare this output current waveform with that shown in Figure 37-19. Note that crossover distortion does not occur when a small forward bias is applied.

Q12. What is crossover distortion?

Q13. How does the application of a small forward bias eliminate crossover distortion?

COMPLEMENTARY SYMMETRY

Junction transistors are available as PNP and NPN types. The direction of electron current flow in the terminal leads of the one type of transistor is opposite to that in the corresponding terminal lead of the other type transistor.

If the two types of transistors are connected in a single stage (Figure 37-23), the dc electron current path (indicated by arrows) in the output circuit is completed through the collector-emitter junctions of the transistors. When connected in this manner, the circuit is referred to as a complementary symmetry circuit. The complementary symmetry circuit provides all the advantages of conventional push-pull amplifiers without the need for a phase-inverter driver stage, or for a center-tapped input transformer.

37-12. Complementary Symmetry Circuit

Figure 37-22 shows two transistors in a complementary symmetry connection. Transistor

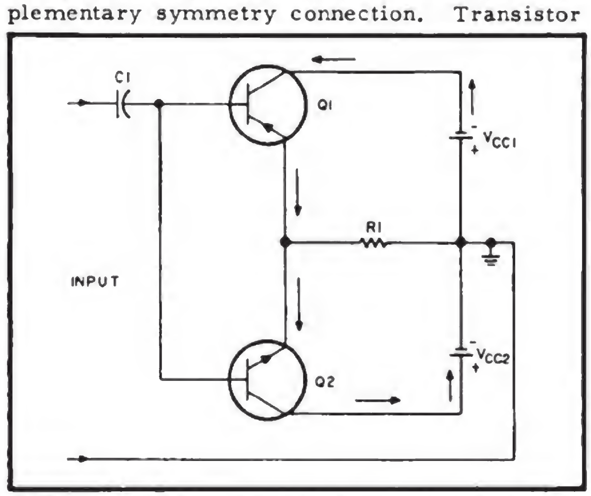

Figure 37-23 - Zero bias complementary symmetry circuit.

Q_1 is a PNP transistor and transistor Q_2 is an NPN transistor. A negative going input signal forward biases transistor Q_1 and causes it to conduct. A positive going input signal forward biases transistor Q_2 and causes it to conduct. As one transistor conducts, the other is nonconducting, because the signal that forward biases one transistor, reverse biases the other transistor.

The resultant action in the output circuit can be understood by considering the circuit of Figure 37-24. This is a simplified version of the output circuit. The internal emitter-collector circuit of transistor Q_1 is represented by variable resistor R_1 and that of transistor Q_2 by variable resistor R_2. With no input signal and class B operation (zero emitter-base bias), the variable arms of the variable resistors can be considered to be in the OFF positions. No current flows through the transistors nor through load resistor R_L. As the incoming signal goes positive, transistor Q_2 conducts and transistor Q_1 remains nonconducting. Variable resistor R_1 remains in the OFF position. The variable arm of resistor R_2 moves toward point 3 and current passes through the series circuit consisting of battery V_{CC2}, resistor R_L and variable resistor R_2. The amount of current flow depends upon the magnitude of the incoming signal, the variable arm moving toward point 3 for increasing forward bias and toward point 4 for decreasing forward

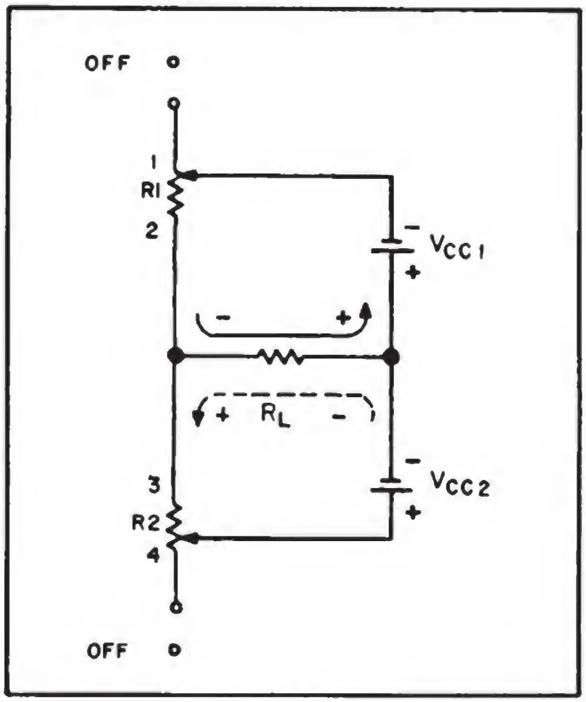

Figure 37-24 - Simplified version of output circuit of complementary symmetry circuit.

bias. The current flows in the direction of the dashed arrow, producing a voltage with the indicated polarity. When the input signal goes negative, transistor Q_1 conducts and transistor Q_2 becomes nonconducting. The same action is repeated with variable resistor R_1. Current flows through battery V_{CC1}, variable resistor R_1, and load resistor R_L in the direction shown by the solid-line arrow, and produces a voltage across resistor R_L with the polarity indicated. For class A operation of a complementary symmetry circuit (such as Figure 37-23), a voltage divider network is used to apply forward bias to the two transistors so that collector current is not cut off at any time. In the simplified circuit (Figure 37-24), the variable resistors will not be in the OFF position at any time. DC bias current in the output circuit flows out of the negative terminal of battery V_{CC2} into the positive terminal of battery V_{CC1}, through variable resistor R_1, variable resistor R_2, and into the positive terminal of battery V_{CC2}. No resultant current flows through resistor R_L. Under these conditions, the output circuit can be considered a balanced bridge, the arms of the bridge consisting of resistors R_1 and R_2 and batters V_{CC1} and V_{CC2}. When the input signal goes positive, transistor Q_2 conducts more and transistor Q_1 conducts less. In the simplified circuit, the variable arm of resistor R_1 moves toward point 1 and that of resistor R_2 moves toward point 3. This action results in an unbalanced bridge and resultant current flows through resistor R_L in the direction of the dashed-line arrow, producing a voltage with the indicated polarity. When the input signal goes negative, transistor Q_1 conducts more and transistor Q_2 conducts less. In the simplified circuit, the variable arm of resistor R_1 moves toward point 2 and that of resistor R_2 moves toward point 4. Again the bridge is unbalanced, and electrons flow through resistor R_L in the direction of the solid-line arrow, producing a voltage with the indicated polarity.

In either class B or class A operation no resultant dc current flows through the load resistor or the primary of the output transformer, whichever is used as the load for the amplifier. Advantage of this property can be taken by connecting the voice coil of a loudspeaker directly in place of resistor R_L. The voice coil will not be offset by dc current flow through it and thus distortion will not occur.

Q14. Would a zero bias complementary symmetry circuit suffer from crossover distortion?

37-13. Output Circuits for Audio Power Amplifiers

An output transformer is most often used as

A12. Distortion resulting in the output of a class B, zero bias, push-pull amplifier due to the non-linear characteristics of the transistors near cut-off.

A13. By moving the operating point away from the non-linear portion of the curve around cut-off.

A14. Yes. At the point where the input signal passes through zero, both transistors would be operating on the non-linear portions of their transfer characteristic curves (near cut-off).

---

the output coupling device for an audio power amplifier. Output transformers used in transistorized superheterodyne receivers vary in size and weight from 1/2 an ounce up to a pound, depending upon power output and other design requirements.

The audio output transformer serves two primary functions; that of impedance matching and of providing dc isolation for the speaker voice coil. Since the impedance of most speakers is lower than the power amplifier output impedance, output transformers are usually of the step down type.

For impedance matching, it should be remembered that the ratio of primary impedance (Z_P) to secondary impedance (Z_S) equals the square of the primary-to-secondary turns ratio:

$$\frac{Z_P}{Z_S} = \left(\frac{N_P}{N_S}\right)^2$$

or:

$$\frac{N_P}{N_S} = \sqrt{\frac{Z_P}{Z_S}}$$

where:

N_p and N_S = the number of primary and secondary turns respectively.

Z_p and Z_S = the impedance of the primary and secondary respectively.

By knowing the impedance ratio, a transformation can be selected to provide the required impedance match; or, given a particular transformers turns ratio, the impedance ratio it will match can be found.

It should be noted that in a single-ended power amplifier, current always flows in one direction through the primary winding. Over a period of time this current will tend to magnetize the transformer core. Such magnetization shifts transformer operation towards the saturation region of the core and interfers with field build-up and collapse around the transformer windings. This results in a distorted output for large amplitude signals. One method of overcoming this problem is to periodically reverse the transformer primary connections.

The advantage of push-pull operation, in this respect, is the dc current flows in opposite directions in each half of the primary winding, thus, the resultant residual magnetic effects in the core are cancelled.

As was mentioned during the discussion of complimentary symmetry circuits, the output transformer can be completely eliminated (under certain conditions) and the power transistor stage connected directly to the speaker voice coil. In other words, if properly designed the speaker voice coil can take the place of the output transformer primary in some applications.

Q15. What value of load impedance would have to be used to match the 32 k ohm output impedance of an audio amplifier if the only available output transformer had a turns ratio of approximately 45:1 (44. 72:1 to be exact)?

EXERCISE 37

1. How may transistor audio amplifiers be connected to high source resistances without the use of a transformer? List the advantages and/or disadvantages of the methods used.
2. Under what class of operation must single-ended audio amplifiers be operated? Why?
3. Why is RC coupling less efficient than transformer coupling?
4. What limits the frequency response when RC coupling is used (high and low)?
5. What is the function of a volume control? List the requirements of a volume control.
6. Sketch, and explain, a usable volume control circuit.
7. What is the primary function of a power amplifier?
8. Define the terms "small signal", "large signal", and power amplifier.
9. What component is generally used to couple the output of a power amplifier to its load? Why?
10. In what manner does the output transformer of a push-pull amplifier differ from one used with a single ended amplifier?
11. Why is it desirable to have the transfer characteristic curves exhibit approximately straight lines throughout the signal voltage range?
12. What is a phase inverter? What is its purpose?
13. Explain the operation of two types of phase inverters.
14. What is meant by a constant collector power dissipation curve?
15. Describe the operation of a single-ended class A power amplifier.
16. What are the advantages of a two stage phase inverter over other types?
17. Describe a method for prevention of cross-over distortion in a class B push-pull amplifier.
18. Why is a bypass capacitor NOT used in the bias network of a low bias class B push-pull amplifier.
19. What is the result of second harmonics being present in the output signal?
20. Explain the term "second harmonics" and cite an example.
21. What circuit is best suited to the elimination of second harmonics?
22. Describe the principles of operation of a complementary symmetry, push-pull amplifier.
23. What are the advantages of a complementary push-pull amplifier over other types of push-pull amplifiers?
24. Why are most output transformers, used in power amplifiers, of the step down type?
25. Why would dc flowing through a speaker voice coil produce distortion?

A15. Approximately 16 ohms.

CHAPTER 38

ELECTRON TUBE RECEIVERS

Many of the principles, circuits, and components discussed in previous chapters are directly applicable to radio receivers. A basic knowledge of them is therefore assumed in treating the material included in this chapter.

At the radio transmitter the carrier frequency is modulated by the desired signal, which may consist of coded characters, voice, music, or other types of signals. AMPLITUDE MODULATION (AM) occurs if the signals cause the amplitude of the output to vary. FREQUENCY MODULATION (FM) occurs if the signals cause the frequency of the carrier, or center frequency, to vary. Although there are other types of modulation, only AM receivers will be treated in this chapter.

The RF carrier wave with the modulating signal impressed upon it is transmitted through space as an electromagnetic wave to the antenna of the receiver. As the wave passes across the receiving antenna, small ac voltages are induced in the antenna. These voltages are coupled into the receiver via the antenna coupling coil. The function of the receiver is to select the desired carrier frequency from those present in the antenna circuit and to amplify the small ac signal voltage. The receiver then removes the carrier by the process of detection (rectification and the removal of the RF component) and amplifies the resultant audio signal to the proper magnitude to operate the loudspeaker or earphones.

Two major types of radio receivers are reviewed in this chapter—the TUNED-RADIO-FREQUENCY (TRF) receiver, and the SUPERHETERODYNE receiver.

TRF RECEIVERS

The tuned-radio-frequency receiver, generally known as the TRF receiver, consists of one or more RF stages, a detector stage, one or more AF stages, a reproducer, and the necessary power supply. A block diagram of a TRF receiver is shown in Figure 38-1. The waveforms that appear in the respective sections of the receiver are shown below the block diagram.

The amplitude of the AM signal at the input of the receiver is relatively small because it has been attenuated in the space between the transmitter and the receiver. It is composed of the carrier frequency and the modulation envelope. The RF amplifier stages amplify the waveform, but they do not change its basic shape if the circuits are operating properly. The detector rectifies and removes the RF component of the signal. The output of the detector is a weak signal made up only of the modulation component, or envelope, of the incoming signal. The AF amplifier stages following the detector increase the amplitude of the AF signal to a value sufficient to operate the loudspeaker or earphones.

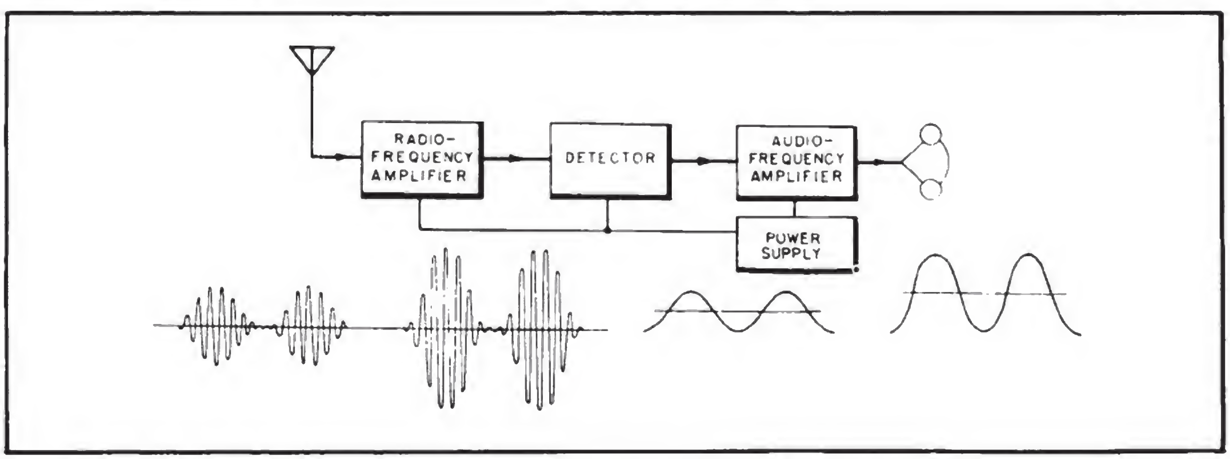

Figure 38-1 - Block diagram of a TRF receiver and waveforms.

38-1. RF Section

The ANTENNA-GROUND SYSTEM serves to introduce the desired signal into the first RF amplifier stage via the antenna coupling transformer. For best reception the resistance of the antenna-ground system should be low. The antenna should also be of the proper length for the band of frequencies to be received; and the antenna impedance should match the input impedance of the receiver. The gain of most commercial receivers, however, is generally sufficient to make these values non-critical.

The RF AMPLIFIERS in the TRF receiver have tunable tanks in the grid circuits. Thus, the receiver may be tuned so that only one RF signal within its tuning range is selected for amplification. When the tank is tuned to the desired frequency, it resonates and produces a relatively large circulating current. The grid of the RF amplifier then receives a relatively large signal voltage at the resonant frequency, and minimum signals at other frequencies.

The relative ability of a receiver to select one particular frequency and to reject all others is called the SELECTIVITY of the receiver. The relative ability of the receiver to amplify small signal voltages is called the SENSITIVITY of the receiver. Both of these values may be improved by increasing the number of RF stages. When this is done, the tuning capacitors in the grid tank circuits are usually ganged on the same shaft and trimmers are added in parallel with each capacitor to make the stages track at the same frequency. In addition, the outer plates of the rotor sections of the capacitors are sometimes slotted to enable more precise alignment throughout the tuning range.

Tetrodes or pentodes are generally used in RF amplifiers because, unlike triodes, they do not usually require neutralization. They also have higher gain than triodes.

A typical RF amplifier stage employing a pentode is shown in Figure 38-2. Tuned circuit L_2C_1 is inductively coupled to L_1, the antenna coil. R_1 and C_3 provide operating bias for the tube. C_4 and R_2 are the screen bypass capacitor and dropping resistor, respectively. The tuned circuit, L_4C_6, couples the following stage inductively to L_3. Both transformers are of the air-core type. The dotted lines indicate mechanical ganging of C_1 and C_6 on the same shaft. The tuning capacitor in the next stage is also ganged on the same shaft.

If it is desired to cover more than one frequency range, additional coils having the proper inductance are used. They are sometimes of the plug-in variety, but more generally they are mounted on the receiver and their leads are connected to a multicontact rotary switch.

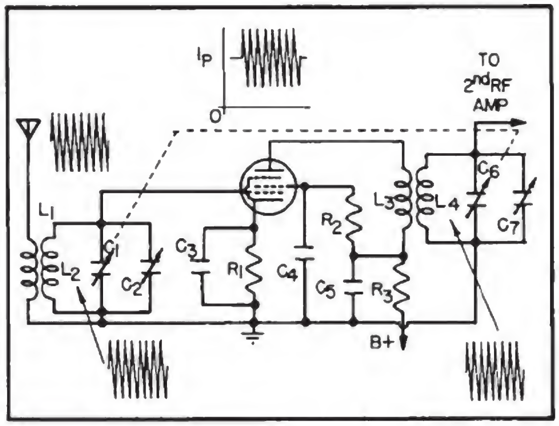

Figure 38-2 - RF amplifier stage.

The latter method is preferable for BAND SWITCHING because the desired band can be selected simply by turning the switch. The same tuning capacitor is used for each band. However, when band switching is employed, the trimmers are connected across the individual tuning inductors and not across the main tuning capacitors.

Decoupling circuits are designed for both RF and AF amplifiers to counteract feedback. Thus, in the RF amplifier in Figure 38-2, C_5 and R_3 make up the decoupling circuit. R_3 offers a high impedance to the signal current, but C_5 offers a low impedance. Consequently the signal current is shunted to ground around the "B" supply. Since R_3 also offers a high resistance to dc current, it may be replaced by a choke coil having a high impedance only to the signal current. Each stage is similarly equipped with a decoupling circuit.

A mechanical or an electrical bandspread may be used as an aid in separating stations that are crowded together on the tuning dial. MECHANICAL BANDSPREAD is simply a micrometer arrangement to reduce the motion of the capacitor rotor as the tuning knob is turned. When ELECTRICAL BANDSPREAD is used a small variable capacitor is connected in parallel with the tuning capacitor. Because of its small size, this variable capacitor may be moved a considerable amount before it causes an appreciable change in the frequency of the tuned circuit. If the tuning capacitors are ganged, the bandspread capacitors are also ganged.

38-2. Detector

The process of removing the intelligence component of the modulated waveform from the RF carrier is called DETECTION or DEMODULATION. In the AM system the audio or

intelligence component causes both the positive and the negative half cycles of the RF wave to vary in amplitude. The function of the detector is to rectify the modulated signal. A suitable filter eliminates the remaining RF pulses and passes the audio component on to the AF amplifiers.

Details of the various methods of detection will be discussed in a subsequent paragraph. Each of the several methods that might be used in the TRF receiver have certain inherent weaknesses. For example, the diode detector requires several stages of amplification ahead of the detector. It loads its tuned input circuit, and therefore the sensitivity and selectivity of the circuit are reduced. However, it can handle strong signals without overloading, and its linearity is good.

The grid-leak detector is sensitive (and therefore requires fewer stages of amplification), but it has poor linearity and selectivity and it may be overloaded on strong signals.

The circuit shown in Figure 38-3 employs plate detection. It has medium sensitivity and the ability to handle strong signals without overloading. The selectivity of this circuit is excellent, but because the i_p-e_g graph is curved near the cutoff point (where the plate detector operates) some distortion in the output cannot be avoided.

In Figure 38-3, the tube is biased nearly to cutoff by the average plate current that flows through R_1. This average value increases as the signal strength increases. On positive half cycles of the incoming signals the plate current varies with the amplitude of the modulating wave and produces the desired AF output voltage. On negative half cycles no appreciable plate current flows. Between positive half cycles the bias voltage is held constant across R_1 by the action of C_3, because the time constant of R_1C_3 is long compared with the time for the lowest AF cycle.

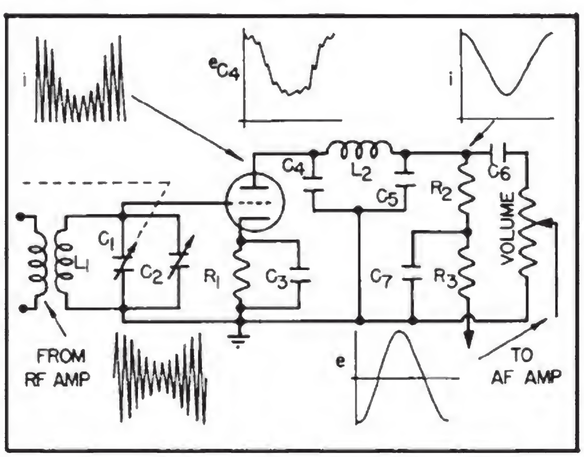

Figure 38-3 - Plate detector circuit.

The RF pulses are filtered out by means of a low-pass filter (consisting of C_4, L_2, and C_5), which rejects the RF component and passes the AF component. C_6 couples the AF component to the first audio amplifier. R_2 is the plate load resistor, and the combination R_3C_7 makes up the decoupling circuit.

38-3. AF Section

The function of the AF section of a receiver is to further amplify the audio signal, which is commonly fed via the volume control to the grid of the first audio amplifier tube. In most cases the amount of amplification that is necessary depends on the type of reproducer used. If the reproducer consists of earphones, only one stage of amplification may be necessary. If the reproducer is a large speaker or other mechanical device requiring a large amount of power, several stages may be necessary. In most receivers the last AF stage is operated as a power amplifier.

A necessary part of the AF section is some means of manual control of the output signal level of the receiver.

A MANUAL VOLUME CONTROL may be employed in a number of receiver circuits. Normally this control varies the amplitude of the signal applied to the grid of an amplifier tube, as shown in Figure 38-3. Increasing the resistance between ground and the sliding contact increases the amplitude of the signal applied to the grid of the driven stage.

An AF OUTPUT STAGE is shown in Figure 38-4. C_1 couples the first AF amplifier to the output stage, and R_1 is the grid coupling resistor. R_2 and C_2 provide a steady bias. Because of the low frequencies involved, C_2 should have a larger value of capacitance than similar bypass capacitors in the RF section. C4 is the plate-bypass, or decoupling capacitor. C3 has

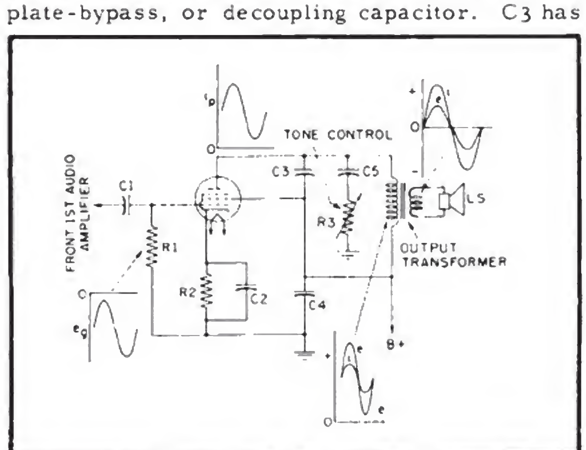

Figure 38-4 - Audio amplifier output stage.

a small value of capacitance and bypasses some of the higher frequencies around the output transformer, thus emphasizing the bass. The impedance of the output-transformer primary commonly represents a compromise between maximum power transfer and minimum distortion. The impedance of the secondary is chosen to match the impedance of the voice coil. Some secondaries have taps on the windings to permit an impedance match to a variety of voice-coil impedances.

TONE CONTROL may be used in communications receivers. The purpose of tone control is to emphasize either the low or the high frequencies by shunting the undesired frequencies around the remainder of the circuit components in the audio section. A simple tone-control circuit, such as the series capacitor C5 and variable resistor R3 combination shown in Figure 38-4, may be connected between plate and ground or between grid and ground in any of the audio stages of a receiver. In this figure it is connected between plate and ground. The value of the series capacitor is such that it will bypass to ground the high-frequency components. The amount of high-frequency energy removed by the tone-control circuit is determined by the setting of the variable-resistor control arm. When the resistance is low, the high frequencies are attenuated; when it is high they appear in the output.

Feedback voltage from output to input is sometimes developed across the impedance of the common power supply. For frequencies within the usable audio range, this impedance is sufficiently low so that insufficient feedback is obtained to cause oscillation. However, for extremely low frequencies, the capacitors in the power supply will sometimes have enough impedance to cause oscillation.

MOTORBOATING IN AUDIO STAGES.—When two or more audio amplifier stages are supplied from a common B supply, feedback occurs as a result of common coupling between the plate circuits, and some method of decoupling must be employed. The coupling consists of the internal impedance of the source of plate voltage. The feedback may either increase or decrease the amplification depending on the phase relation between the input voltage and the feedback voltage. In a multistage amplifier the greatest transfer of feedback energy occurs between the final and first stages because of the high amplification through the multistage amplifier.

The effects of feedback are important if the feedback voltage coupled into the plate circuit of the first stage is appreciable compared to the signal voltage that would be developed if feedback did not exist. For example, a three-stage resistance-coupled amplifier may develop a feedback voltage (coupled via the B supply into the plate circuit of stage 1) which is in phase with the signal voltage of stage 1 and hence may cause oscillations to be set up. In audio amplifiers having high gain and a good low-frequency response this regeneration causes a low-frequency oscillation known as "motorboating" because of the "putt-putt" sound in the speaker.

Design engineers usually decouple plate circuits by adding a series resistor to the input stage, between its plate load and B+, and bypassing that resistor to ground. The appearance of motorboating reveals the need of replacing either the decoupling resistor or its bypass capacitor.

38-4. Circuit of the TRF Receiver

The complete circuit of a TRF radio receiver operated from an ac power supply is shown in Figure 38-5. The receiver uses two pentodes in the RF section, one triode operated as a plate detector, and two pentode AF amplifier stages that feed the loudspeaker.

From previous discussions, the various circuits may be identified and the signal may be traced from the antenna-ground system to the loudspeaker. The dotted lines indicate that the three main tuning capacitors are ganged on a single shaft. Across each of the main tuning capacitors is connected a trimmer capacitor to enable circuit alignment. The ground circuit and the various decoupling circuits may be readily identified. The power supply voltage is obtained from a conventional full-wave rectifier. Rectifier and tube filament currents are obtained from two low-voltage windings on the power transformer.

38-5. Characteristics of the TRF Receiver

The principal disadvantage of the TRF receiver is that its selectivity, or its ability to separate signals, does not remain constant over its tuning range. As the set is tuned from the low-frequency end of its tuning range to the high frequency end, its selectivity decreases.

Also, the amplification, or gain, of a TRF receiver is not constant over the tuning range of the receiver. The gain depends on RF transformer gain, which increases with frequency. In order to improve the gain at the low-frequency end of the band, RF transformers employing high-impedance (untuned) primaries are designed so that the primary inductance will resonate with the primary distributed capacitance at some frequency slightly below the low end of the tunable band. Thus, the gain is good at the low end of the band because of the resonant buildup of primary current. The near-resonant condition of the primary at the low end more than offsets the effect of reduced transformer

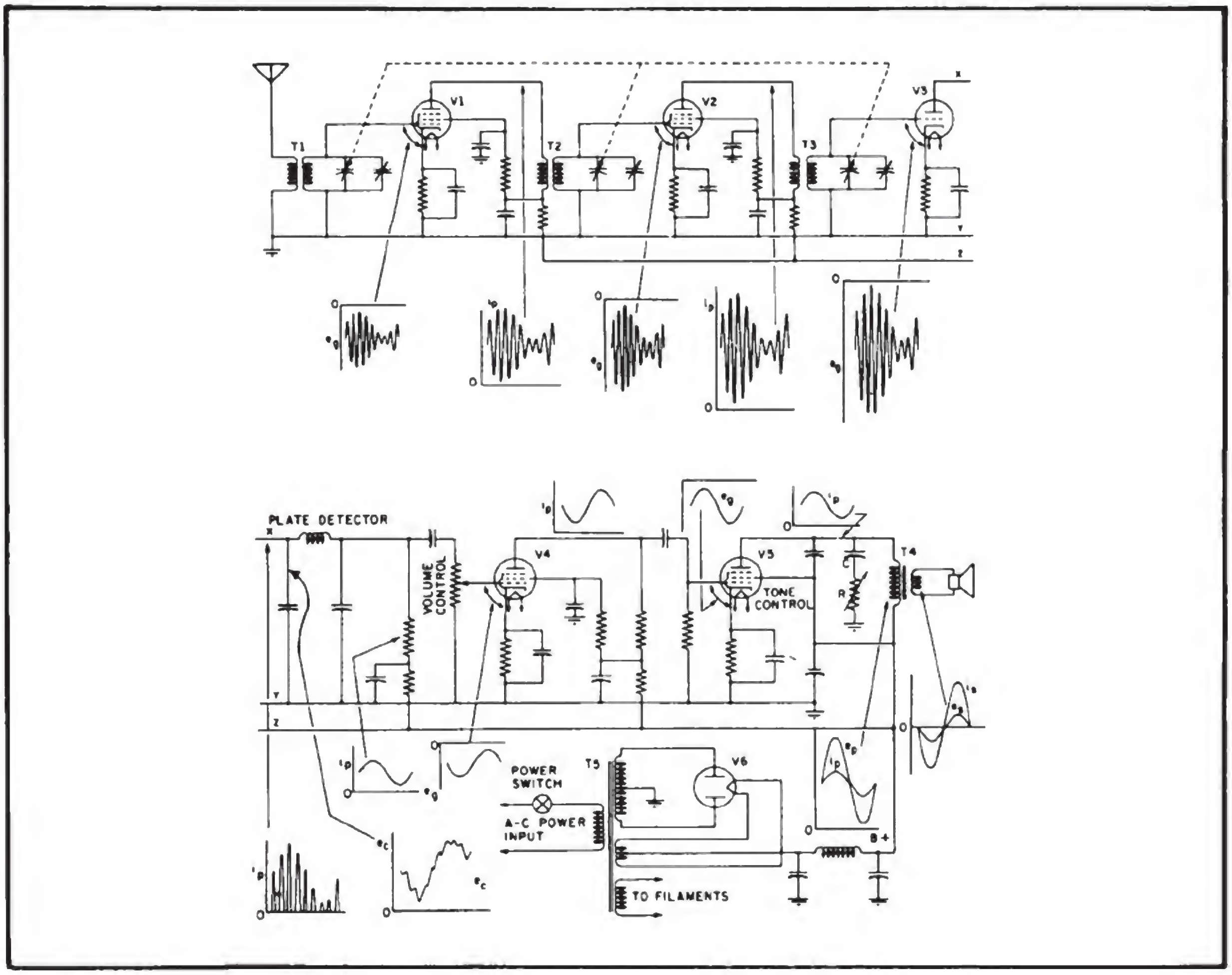

Figure 38-5 - Circuit of a TRF receiver.

action. However, the shunting action of the primary distributed capacitance lowers the gain at the high-frequency end of the band. To make up for the resultant poor gain at the high end of the band, a small capacitor is connected between the plate and grid leads of adjacent RF stages to supplement the transformer coupling. At the low end of the band the capacitive coupling is negligible.

The superheterodyne receiver has been developed to overcome many of the disadvantages of the TRF receiver.

SUPERHETERODYNE RECEIVERS

The essential difference between the TRF receiver and the superheterodyne receiver is that in the former the RF amplifiers preceding the detector are tunable over a band of frequencies; whereas in the latter the corresponding amplifiers are tuned to one fixed frequency called the INTERMEDIATE FREQUENCY (IF). The principle of frequency conversion by heterodyne action is here employed to convert any desired station frequency within the receiver range to this intermediate frequency. Thus an incoming signal is converted to the fixed intermediate frequency before detecting the audio signal component, and the IF amplifier operates under uniformly optimum conditions throughout the receiver range. The IF circuits thus may be made uniformly selective, uniformly high in voltage gain, and uniformly of satisfactory bandwidth to contain all of the desired sideband components associated with the amplitude modulated carrier.

The block diagram of a typical superheterodyne receiver is shown in Figure 38-6. Below corresponding sections of the receiver are shown the waveforms of the signal at that point. The RF signal from the antenna passes first

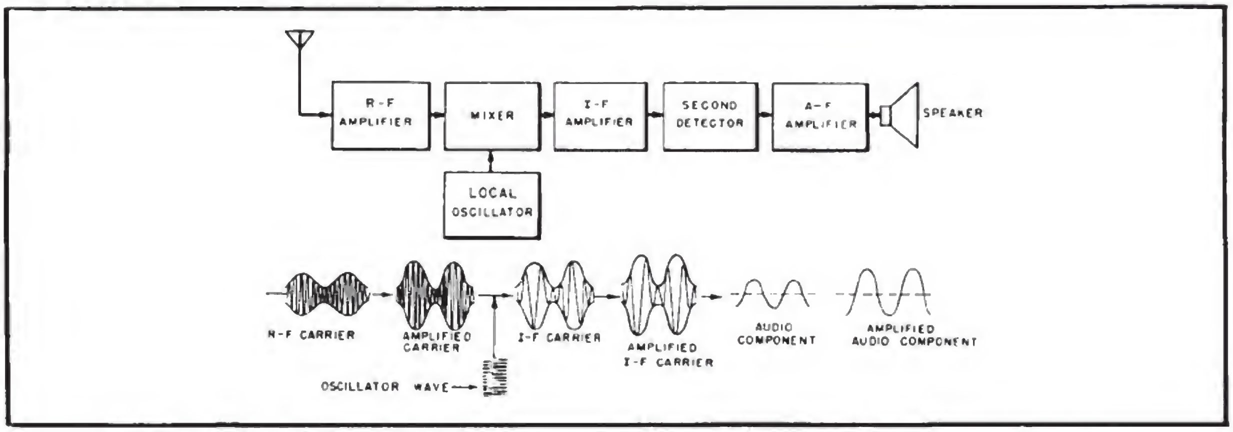

Figure 38-6 - Block diagram of a superheterodyne receiver and waveforms.

through an RF amplifier (preselector) where the amplitude of the signal is increased. A locally generated unmodulated RF signal of constant amplitude is then mixed with the carrier frequency in the mixer stage. The mixing or heterodyning of these two frequencies produces an intermediate frequency signal which contains all of the modulation characteristics of the original signal. The intermediate frequency is equal to the difference between the station frequency and the oscillator frequency associated with the heterodyne mixer. The intermediate frequency is then amplified in one or more stages called INTERMEDIATE FREQUENCY (IF) AMPLIFIERS and fed to a conventional detector for recovery of the audio signal.

The detected signal is amplified in the AF section and then fed to a headset or loudspeaker. The detector, the AF section, and the reproducer of a superheterodyne receiver are basically the same as those in a TRF set, except that diode detection is generally used in the superheterodyne receiver. Automatic volume control or automatic gain control also is commonly employed in the superheterodyne receiver.

Q1. Why would a triode, used as an RF amplifier require neutralization?

Q2. What are the main disadvantages of the TRF receiver?

38-6. RF Amplifier

If an RF amplifier is used ahead of the mixer stage of a superheterodyne receiver it is generally of conventional design. Besides amplifying the RF signal, the RF amplifier has other important functions. For example, it isolates the local oscillator from the antenna-ground system. If the antenna were connected directly to the mixer stage, a part of the local oscillator signal might be radiated into space. This signal could be picked up by a sensitive direction finder on any enemy ship. For this reason and others, Navy superheterodyne receivers are provided with at least one RF amplifier stage.

Also, if the mixer stage were connected directly to the antenna, unwanted signals, called IMAGES, might be received, because the mixer stage produces the intermediate frequency by heterodyning two signals whose frequency difference equals the intermediate frequency. (The heterodyne principle is treated later in this chapter.)

The image frequency always differs from the desired station frequency by twice the intermediate frequency—Image frequency = station frequency ± (2 X intermediate frequency). The image frequency is higher than the station frequency if the local oscillator frequency tracks (operates) above the station frequency (Figure 38-7, A). The image frequency is lower than the station frequency if the local oscillator tracks below the station frequency (Figure 38-7, B). The latter arrangement is generally used for the higher frequency bands, and the former, for the lower frequency bands.

For example, if such a receiver having an intermediate frequency of 455 kc is tuned to receive a station frequency of 1500 kc (Figure 38-7, A), and the local oscillator has a frequency of 1955 kc, the output of the i-f amplifier may contain two interfering signals—one from the 1500-kc station and the other from an image station of 2410 kc (1500 + 2 X 455 = 2410 kc). The same receiver tuned near the low end of the band to a 590-kc station has a local oscillator frequency of 1045 kc. The output of the IF amplifier contains the station signal (1045-590 = 455 kc) and an image signal (1500-1045 = 455 kc). Thus the 1500-kc signal is an image

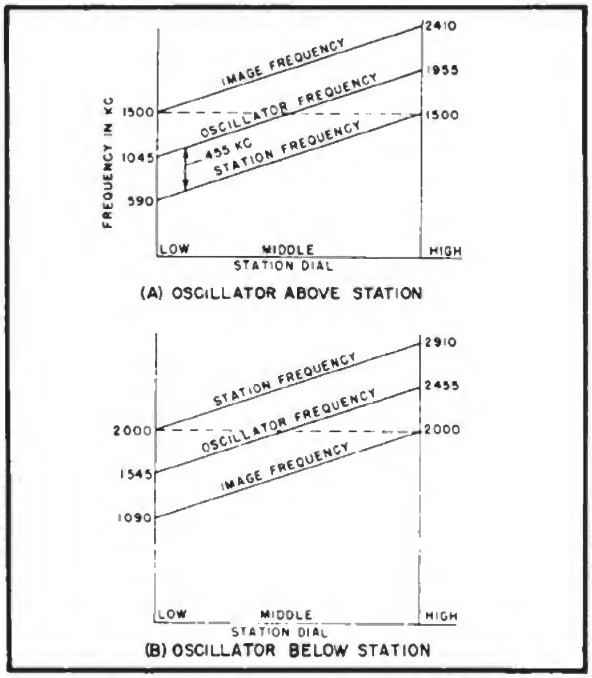

Figure 38-7 - Relation of image frequency to station frequency in a superheterodyne receiver.

heard simultaneously with the 590-kc station signal.

It may also be possible for ANY two signals having sufficient strength, and separated by the intermediate frequency to produce unwanted signals in the reproducer. The selectivity of the preselector tends to reduce the strength of these images and unwanted signals.

The ratio of the amplitude of the desired station signal to that of the image is called the IMAGE REJECTION RATIO and is an important characteristic of a superheterodyne receiver. Better superheterodyne receivers are therefore equipped with one or more preselector stages, a typical example of which is shown in Figure 38-9.

The amplification of a tube may be controlled by varying the bias voltage applied to the grid, but normally the range of this control is limited by the cutoff bias and the permissible distortion. In receivers employing automatic volume control (AVC) in the RF amplifier section, the amplification is varied over a wide range so that strong or weak signals may be accommodated. To permit this increased range of volume control, the variable-mu tube was developed. This tube is also known as the supercontrol or remote cutoff type.

The only difference in construction between variable-mu tubes and normal, or sharp cutoff tubes in in the spacing between the turns of the control grid. In sharp cutoff tubes the turns of the grid wires are equally spaced, while in remote cutoff types the grid turns are closely spaced at the ends and widely spaced in the center. The construction of variable-mu tubes is shown in Figure 38-8, A.

With a small bias voltage, electrons flow through all the spaces of the grid and the amplification factor is relatively large because of the close spacing of the end turns of the control grid. As the bias is increased, the electron flow is cut off through the narrow spaces at the ends of the grid structure. However, they are still able to pass through the relatively large spaces at the center of the grid. The increased bias causes a decrease in the amplification due to the coarser turns in the central portion of the grid. A much greater value of bias is required to cut off the plate current flow in this type of tube. The remote-cutoff tube is so named because the cutoff bias value is greater than (more remote from) the value required to cut off plate current flow in tubes of evenly spaced turns.

Figure 38-8, B, shows the i_p-e_g curves for both a conventional sharp cutoff tube and a variable-mu or remote-cutoff tube. The cutoff bias for the normal tube is -5 volts, and because the slope is almost constant any change in bias produces little change in amplification. Contrasted

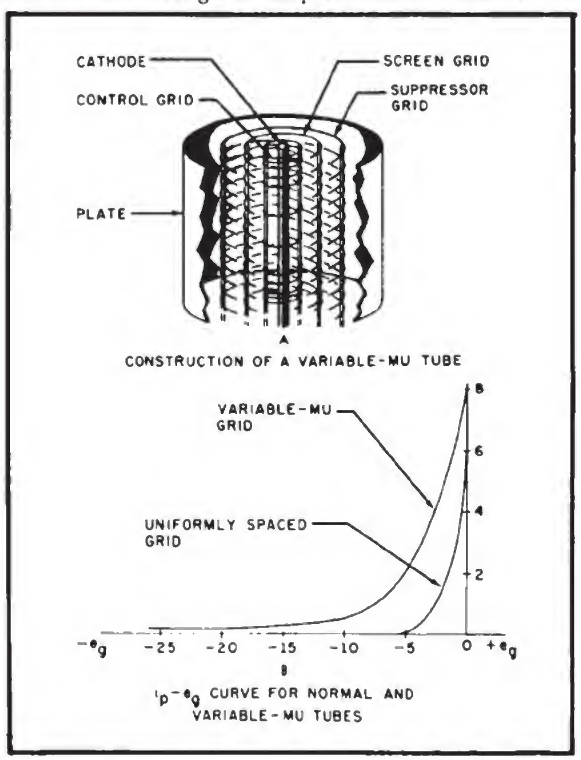

Figure 38-8 - Construction of variable-mu tubes and i_p-e_g curves.

A1. The triode would require neutralization to suppress the effects of feedback, which occur due to the high grid to plate capacity.

A2. The main disadvantages of the TRF receiver lie in the variations in selectivity and gain which occur over the tuning range.

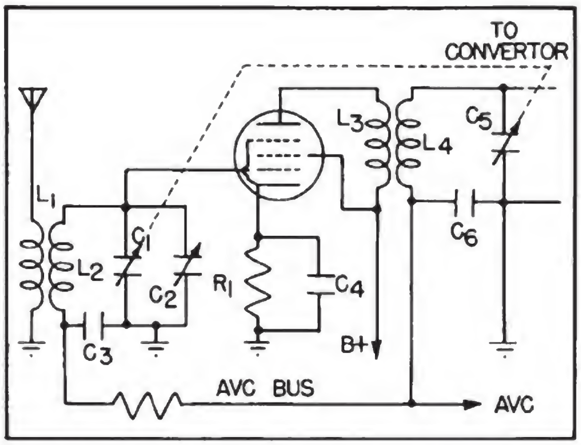

Figure 38-9 - Typical superheterodyne preselector stage.

with this characteristic, the curve for the variable-mu tube has a pronounced change in slope as the grid bias is increased from -10 volts to -15 volts and a small value of plate current is still flowing at a bias of -25 volts. The changing slope of this curve indicates a variation of amplification with bias. Thus, if a variable-mu tube is used with a bias source that varies with the signal strength, (AVC bus Figure 38-9), the output signal can be made substantially independent of the input signal strength.

The preselector stage employs a variable-mu tube and cathode bias. L_1 is the antenna coil. L_2 and C_1 make up the tuned input circuit, and C_2 is the trimmer used for alignment purposes. The dotted line indicates ganged tuning capacitors. Usually these are the tuning capacitor of the mixer input tank circuit and the local oscillator tuning capacitor. C_3 provides low impedance coupling between the lower end of L_2 and the grounded end of C_2, thus bypassing the decoupling filters in the automatic-volume-control (AVC) circuit. (Automatic-volume, or automatic-gain, control is treated later in this chapter.) The RF transformer in the output circuit consists of an untuned high-impedance primary, L_3, and a tuned secondary, L_4, which resonates with tuning capacitor C_5 at the station frequency. RF bypass capacitor C_6 serves a function similar to that of C_3.

38-7. First Detector

The first detector, or frequency-converter, section of a superheterodyne receiver is composed of two parts—the oscillator and the mixer. In many receivers, particularly at broadcast frequencies, the same vacuum tube serves both functions, as in the pentagrid converter shown in Figure 38-10. The operation of the tube may be simplified somewhat if both stages (oscillator and mixer) are considered as exerting two different influences on the stream of electrons from cathode to plate. These electrons are influenced by the oscillator stage (grids, 1, 2, and 4) and also by the station input signal on grid number 3. Thus, coupling between the input signal and the oscillator takes place within the electron stream itself.

There is a tendency for the local oscillator to synchronize with the station frequency signal applied to grid 3. At high frequencies where the two signals have nearly the same frequency, the pentagrid converter is replaced with a mixer tube and a separate oscillator tube. This type of circuit provides frequency stability for the local oscillator.

The oscillator stage employs a typical Hartley circuit in which C_5 and the oscillator coil make up the tuned circuit. C_4 is the trimmer capacitor which is used for alignment (tracking) purposes. C_3 and R_2 provide grid-leak bias for the oscillator section of the tube. Grid 1 is the oscillator grid, and grids 2 and 4 serve as the oscillator plate. Grids 2 and 4 are connected together and also serve as a shield for the signal input grid, 3.

Grid 3 has a variable-mu characteristic, and serves as both an amplifier and a mixer grid. The tuned input is made up of L_1 and

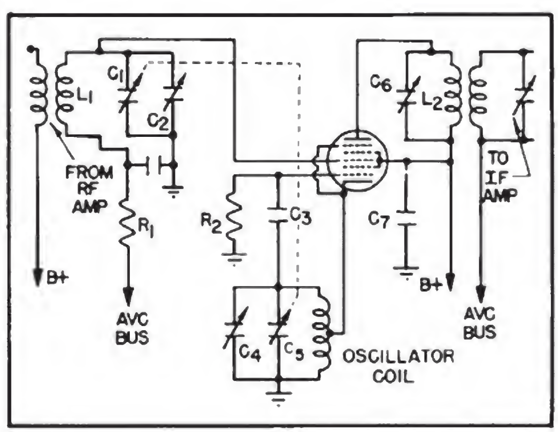

Figure 38-10 - First detector employing a pentagrid converter.

C1, with the parallel trimmer C_2. The dotted lines drawn through C_1 and C_5 indicate that both of these capacitors are ganged on the same shaft (in this example with the preselector tuning capacitor). The plate circuit contains the station frequency and the oscillator frequency signals both of which are bypassed to ground through the low reactance of C_6 and C_7. The heterodyne action within the pentagrid converter produces additional frequency components in the plate circuit, one of which is the difference frequency between the oscillator and the station frequency. The difference frequency is the intermediate frequency and is developed across C_6 and L_2. This signal is coupled to the first IF amplifier grid through the desired band-pass coupling which is wide enough to include the sideband components associated with the amplitude-modulated signal applied to grid 3 of the pentagrid converter.

The conversion gain in a pentagrid converter is,

$$\mu = V_d S_c, \tag{38-1}$$

where V_d is the ac plate resistance with the station RF carrier applied, and S_c is the conversion transconductance (30% to 40% of the g_m of the pentode amplifier). Conversion gain is the change in plate voltage at the intermediate frequency divided by the change in grid voltage at the RF station frequency for equal changes in plate current at the intermediate frequency. Expressed as a formula,

$$\text{conversion gain} = \frac{\text{IF output volts}}{\text{RF input volts}}. \tag{38-2}$$

The conversion gain of a typical pentagrid converter used in broadcast receivers ranges between 30 and 80.

38-8. Heterodyne Principle

The production of audible beat notes is a phenomenon that is easily demonstrated. For example, if two adjoining piano keys are struck simultaneously, a tone will be produced that rises and decreases in intensity at regular intervals. This action results from the fact that the rarefactions and compressions produced by the vibrating strings will gradually approach a condition in which they reinforce each other at regular intervals of time with an accompanying increase in the intensity of the sound. Likewise at equal intervals of time, the compressions and rarefactions gradually approach a condition in which they counteract each other, and the intensity is periodically reduced.

This addition and subtraction of the intensities at regular intervals produces BEAT FREQUENCIES. The number of beats produced per second is equal to the difference between the two frequencies.

The production of beats in a superheterodyne receiver is somewhat analogous to the action of the piano, except that with the receiver the process is electrical and the frequencies are much higher. Figure 38-11 indicates graphically how the beat frequency (intermediate frequency) is produced when signals of two different frequencies are combined in the mixer tube. The resultant envelope varies in amplitude at the difference frequency, as indicated by the dotted lines.

In this example, one voltage, e_s, has a frequency of 8 cycles per second and the other voltage, e_o, has a frequency of 10 cycles per second. Initially, the amplitudes of the two voltages add at instant A, but at instant B the relative phase of e_o has advanced enough to oppose e_s, and the amplitude of the resultant envelope is reduced to a value dependent upon e_s. At instant C the relative phase of e_o has advanced enough to permit the amplitudes to add again. Thus, 1 cycle of amplitude variation of the envelope takes place in the time interval that e_o needs to gain 1 cycle over e_s. From Figure 38-11 it may be seen that e_o gains 2 cycles in the interval A to E. Therefore, the beat or difference frequency is 2 cycles per second. In the superheterodyne receiver the amplitude of the oscillator signal is designed to be greater than that of any received signal.

38-9. IF Amplifier

The IF amplifier is a high-gain circuit commonly employing pentode tubes. This amplifier is tuned to the frequency difference be-

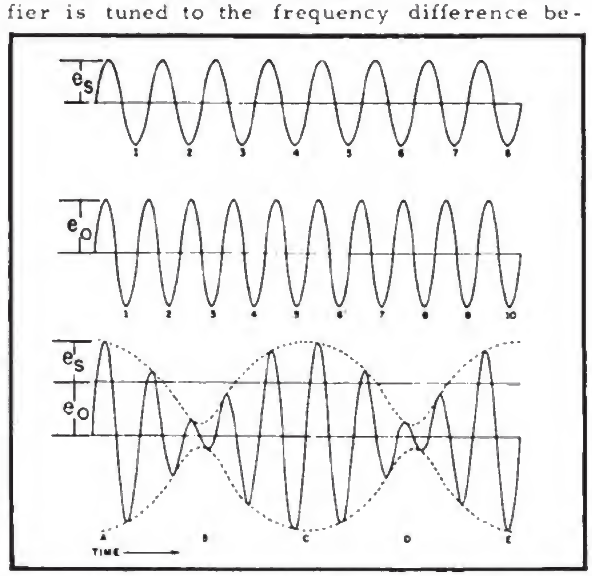

Figure 38-11 - Simplified graphical analysis of the formation of beats.

tween the local oscillator and the incoming RF signal. Pentode tubes are generally employed, with one, two, or three stages, depending on the amount of gain needed. As previously stated, all incoming signals are converted to the same frequency by the frequence converter, and the IF amplifier operates at only one frequency. The tuned circuits, therefore, are permanently adjusted for maximum gain consistent with the desired band pass and frequency response. These stages operate as class-A voltage amplifiers and practically all of the selectivity of the superheterodyne receiver is developed by them.

Figure 38-12 shows the first IF amplifier stage. The minimum bias is established by means of R_1C_1. Automatic volume control may be applied to the grid through the secondary of the preceding coupling transformer.

The output IF transformer, which couples the plate circuit of this stage to the grid circuit of the second IF stage, is tuned by means of capacitors C_2 and C_3. Mica or air-trimmer capacitors may be used. In some instances the capacitors are fixed, and the tuning is accomplished by means of a movable powdered-iron core. This method is called PERMEABILITY tuning. In special cases the secondary only is tuned (single tuned). The coils and capacitors are mounted in small metal cans which serve as shields, and provision is made for adjusting the tuning without removing the shield.

The input IF transformer has a lower coefficient of coupling than the output transformer in some receivers in order to suppress noise from the pentagrid converter. The output IF transformer is slightly overcoupled with double humps appearing at the upper and lower sideband frequencies. The overall response of the stage is essentially flat, and in typical broadcast receivers has a voltage gain of about 200

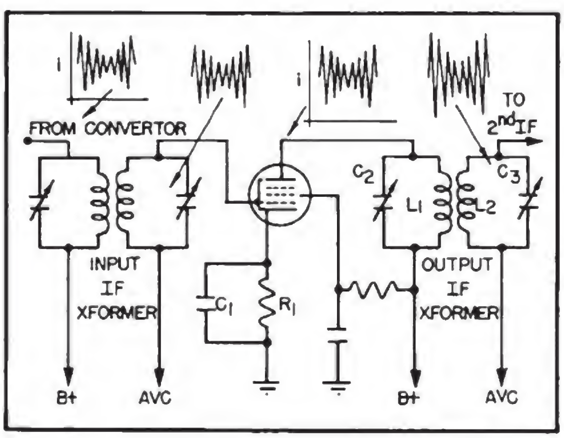

Figure 38-12 - First IF amplifier stage.

with a bandpass of 7 to 10 kc and an IF of about 456 kc.

The chief characteristic of the double-tuned bandpass coupling is that at frequencies slightly above and slightly below the intermediate frequency the impedance coupled into the primary by the presence of the secondary is reactive. This cancels some of the reactance existing in the primary, and the primary current increases. Thus the output voltage of the secondary does not fall off and the response is uniform within the pass band.

38-10. Demodulation of Waves

DEMODULATION, or DETECTION, is the process of recovering the intelligence from a modulated wave. When a radio carrier wave is amplitude-modulated, the intelligence is imposed on the carrier in the form of amplitude variations of the carrier. The demodulator of an amplitude-modulated (AM) wave produces currents or voltages that vary with the amplitude of the wave. Likewise, the frequency-modulation (FM) detector and the phase-modulation (PM) detector change the frequency variations of an FM wave, and the equivalent phase variations of a FM wave into currents or voltages that vary in amplitude with the frequency or phase changes of the carrier.

The instantaneous amplitude, e_o, of the carrier may be represented as

$$e_o = E_o \sin(2\pi f_o t + \theta), \qquad (38\text{-}3)$$

where E_o is the maximum amplitude of the original carrier, f_o the frequency of the carrier, and θ the phase angle (for AM signals, θ may be considered as zero). One or more of the independent variables (those on the right-hand side of the equation) may be made to vary in accordance with the modulating signal to produce a variation in E_o. However, the general practice is to vary only one of the values—E_o (for AM), f_o (for FM), or θ (for PM)—and to prevent any variation in the others.

The detector in the receiver must therefore be designed so that it will be sensitive to the type of modulation used at the transmitter, and insensitive to any other.

Most Navy equipment is designed for amplitude modulation. A clear understanding of the mechanism of AM detection is therefore very important.

AM modulators and demodulators are nonlinear devices. A NONLINEAR DEVICE is one whose current-voltage relation is not a straight line. Because the ratio of current to voltage is not constant, the device has a nonlinear impedance—for example, one of the electron-tube detectors to be considered later—the average out-

put current is the difference between each successive positive and negative swing of the output signal current, as shown in Figure 38-13. The average output (signal component) follows the envelope of the incoming modulated wave more or less closely, depending on the shape of the nonlinear curve. Because the envelope of the incoming AM wave contains the desired audio frequency, a nonlinear device demodulates the AM wave.

For an understanding of the differences in the output frequencies of the various detectors it is necessary to examine the frequencies involved in both modulation and demodulation.

Q3. What is motorboating in an audio amplifier?

Q4. What is the purpose of the small trimmer capacitors in parallel with the tuned circuits of a receiver?

38-11. Comparison of Amplitude Modulation and Demodulation

If at the transmitter an RF carrier and a single-frequency audio-modulating signal of sine waveform are impressed on a LINEAR device, the output waveform from the linear device will contain the same RF and AF signal frequencies. The tuned RF amplifiers in the transmitter will amplify the RF carrier, but will eliminate for all practical purposes the AF component. Under these circumstances only the carrier will be radiated; and it will be ineffective in "carrying" the intelligence component.

A very different result is obtained if an RF carrier and a single frequency audio-modulating signal of sine waveform are impressed on a NONLINEAR device. In this instance distortion is introduced and, as a result, additional frequencies are produced. In addition to the original frequencies, sum-and-difference frequencies are generated, and a zero-frequency, or dc component, is added. The tuned circuits at the transmitter now respond to the carrier and the upper and lower sidebands; but, as before, the AF modulating signal is discriminated against. However, this AF component is replaced, or generated, by the demodulator in the receiver.

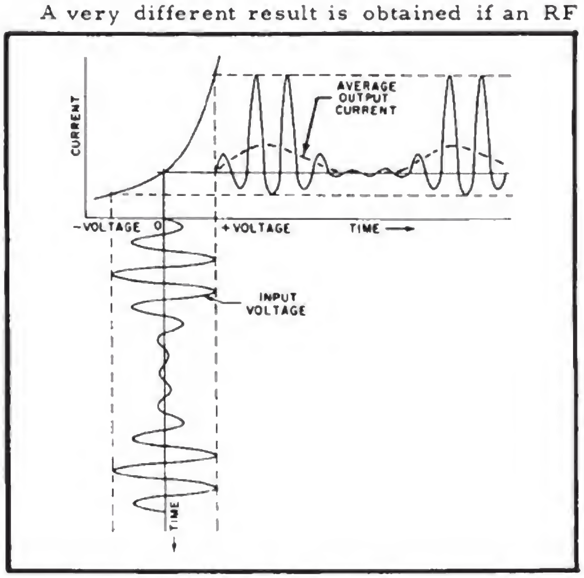

Figure 38-13 - Results of impressing an AM wave on a nonlinear device.

In the receiver the carrier and the two sidebands are impressed on a second nonlinear device called the DEMODULATOR. If the demodulator has an IDEAL nonlinear curve it will distort the incoming waveform (the positive halves of the cycle will be different from the negative halves). Therefore, in addition to the RF carrier and the two sidebands, the SIGNAL FREQUENCY (which is the difference between the upper sideband and the carrier or the difference between the carrier and the lower sideband) and a zero-frequency (or dc component) will be produced. This dc component may be used for automatic volume control.

If the demodulator used in the receiver does not have an ideal nonlinear curve, but has a PRACTICAL realizable curve such as the square-law curve, additional frequencies will be produced. These frequencies will be harmonics of all frequencies present in the input. They are produced because input voltages having larger amplitudes are distorted differently from input voltages having smaller amplitudes. The RF harmonics may be filtered in the output of the demodulator, but the AF harmonics are not easily eliminated.

Thus, modulation and demodulation are essentially the same in that the waveform is distorted in each case and new frequencies are produced.

38-12. Types of AM Detectors

Detectors are classified according to the shape of their current-voltage (characteristic) curve. If the curve is smooth, as in Figure 38-13, the detector is called a SQUARE-LAW DETECTOR. It is called a square-law detector because, for a first approximation, the output voltage is proportional to the square of the effective input voltage.

If the current-voltage curve of the detector is shaped like an obtuse angle, as in Figure 38-14, A the curve is still nonlinear because of the abrupt change in shape at the knee. Because the detector action takes place on the linear portions of the curve on both sides of the

A3. Motorboating is a self-sustained, low frequency oscillation produced in an audio amplifier, due to regenerative feedback occurring between the output and input amplifier stages.

A4. The trimmer capacitors adjust the tracking of the tuned circuits at the high frequency end of the band.

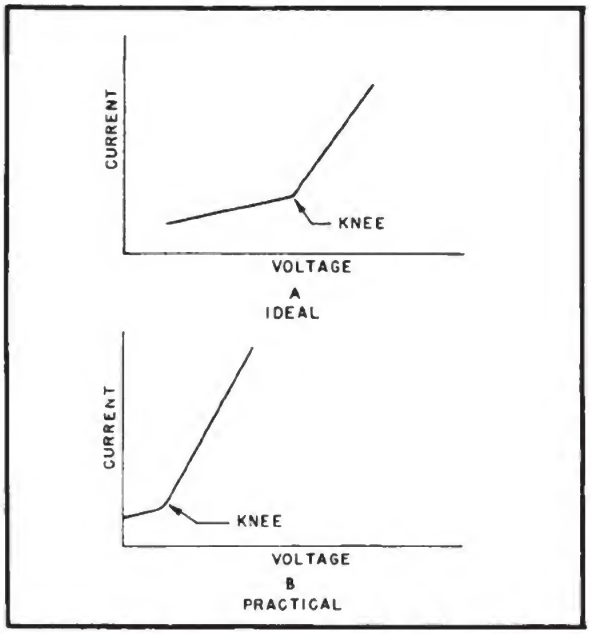

Figure 38-14 - Linear detectors.

knee, this type of detector is called a LINEAR DETECTOR. It is called a linear detector merely to distinguish it from a square-law detector. Both the square-law detector and the linear detector are actually NONLINEAR devices.

The rectified output voltage of a linear detector is proportional to the input voltage. The output of a square-law detector is proportional to the square of the input voltage.

Detectors are also described as power detectors or as weak-signal detectors. If a detector is to handle RF carrier voltages having amplitudes greater than approximately 1 volt it is called a POWER DETECTOR. If the input signal strength is less than this amount the detector is called a WEAK-SIGNAL DETECTOR. Thus, the approximate value of 1 volt is the dividing line between the two detectors. Because a linear detector cannot be obtained with a sharp discontinuity (angle) at the knee, weak signals are always detected on a curved portion of the characteristic, as shown in Figure 38-14, B. Thus, weak-signal detectors are always of the square-law type. Power detectors may be either linear or square-law, depending on the application.

38-13. Diode Detector

The DIODE DETECTOR (Figure 38-15A) is one of the simplest and most widely used detectors, and has nearly an ideal resistance characteristic. Diodes have a point of sharp transition between the conducting (forward) and nonconducting (reverse) directions and therefore make good detectors.

Early radio receivers used a crystal detector that was made of galena, a mineral compound of sulfur and lead. The end of a short length of fine wire touched the surface of the crystal and was held against it by the pressure of a spring. The wire could be adjusted manually. When the operator found a sensitive spot the crystal rectified the signal and the operator heard the signal in his earphones.

Later applications of the crystal detector are found in radar and television. In these applications the crystal diode is of a different form than the older type. It comprises a small crystal with a fine wire point contact bearing on the crystal surface. The components are contained in a sealed ceramic cartridge. The advantage of the crystal diode as a rectifier of high frequency signals is the relatively low shunting capacitance between the crystal and the point contact, compared with the interelectrode capacitance between the plate and cathode of the electron tube diode.

The slight bend in the lower portion of the i_p-e_p characteristic curve for the electron tube diode (Figure 38-15, B) results from contact potential.

Contact potential is the potential difference existing between the surfaces of metals of different electron affinities that are in direct contact or are connected by means of an external circuit. This is true of the metal elements of electron tubes; and the voltages developed may be of the order of 1 volt or more. In the case of the diode, contact potential keeps the plate current from decreasing to zero when the plate voltage approaches zero. The current, however, is very slight, and the characteristic curve is generally shown as zero when the applied voltage is zero. Nevertheless, on low signal voltage, plate current does not increase as rapidly as it does after the contact potential has been exceeded.

Because the diode characteristic is nearly straight on both sides of the knee, the diode detector is a linear detector. However, with weak signals, the output of the detector follows the square law because weak signals force the operation to take place on the lower, curved portion

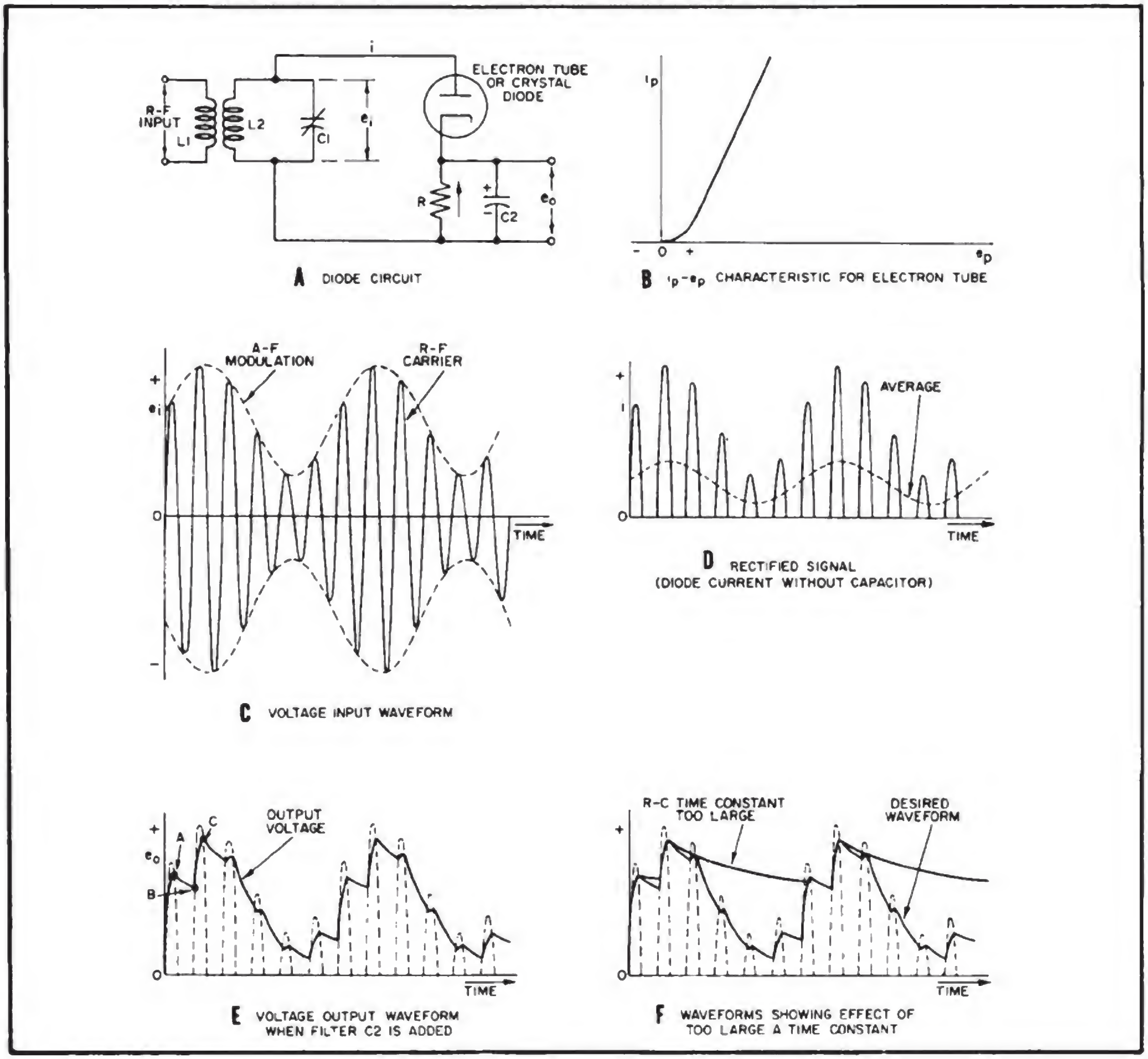

Figure 38-15 - Diode detector and waveforms.

of the characteristic curve (Figure 38-15, B). Because the diode detector normally handles large input signals with minimum distortion, it is classified as a power detector.

The modulated signal voltage (Figure 38-15, C) is developed across the tuned circuit, L_2C_1, of the detector stage. Signal current flows through the diode only when the plate is positive with respect to the cathode—that is, only on the positive half cycles of the RF voltage wave.

The rectified signal flowing through the diode (Figure 38-15, D) actually consists of a series of RF pulses and not a smooth outline or envelope. The average of these pulses, with little or no filtering, does increase and decrease at the AF rate, as shown by the dotted line. Therefore, there is an audio voltage output even if no filtering is employed. However, some stray capacitance exists, and consequently some RF filtering takes place.

If a capacitor (C_2 in Figure 38-15, A) of the proper size is used as a filter, the output voltage of the detector is increased and more nearly follows the envelope. On the first quarter cycle of applied RF voltage, C_2 charges up to nearly the peak value of the RF voltage (point A in Figure 38-15, E). The small voltage drop in the tube prevents C_2 from charging up completely. Then as the applied RF voltage falls below its peak value, some of the charge on C_2 leaks through R, and the voltage across R drops only a slight amount to point B. When the RF

voltage applied to the plate on the next cycle exceeds the potential at which the capacitor holds the cathode (point B), diode current again flows and the capacitor charges up to almost the peak value of the second positive half cycle at point C.

Thus the voltage across the capacitor follows very nearly the peak value of the applied RF voltage and reproduces the AF modulation. The detector output, after rectification and filtering, is a dc voltage that varies at an audio rate, as shown by the solid line in Figure 38-15, E. The curve of the output voltage across the capacitor is shown somewhat jagged. Actually, the RF component of this voltage is negligible and, after amplification, the speech or music originating at the transmitter is faithfully reproduced.

The correct choice of R and C_2 (Figure 38-15, A) in the diode-detector circuit is very important if maximum sensitivity and fidelity are to be obtained. The load resistor, R, and the plate resistance of the diode act as a voltage divider to the received signal. Therefore, the load resistance should be high compared with the plate resistance of the diode so that maximum output voltage will be obtained. The value of C_2 should be such that the RC time constant is long compared with the time of one RF cycle. This is necessary because the capacitor must maintain the voltage across the load resistor during the time when there is no plate current. Also, the RC time constant must be short compared with the time of one AF cycle in order that the capacitor voltage can follow the modulation envelope.

The values of R and C_2 therefore place a limit on the highest modulation (audio) frequency that can be detected. Figure 38-15, F, shows the type of distortion that occurs when the RC time constant is too large. At the higher modulation frequencies the capacitor does not discharge as rapidly as required, and negative peak clipping of the audio signal results.

The efficiency of rectification in a diode is the ratio of the peak voltage appearing across the load to the peak input signal voltage. The efficiency increases with the size of R compared with the diode plate resistance, because R and the diode are in series across the input circuit and their voltages divide in proportion to their resistance. With audio frequencies, a large value of R may be used (of the order of 100,000 ohms), and consequently the efficiency is relatively high (95%). When high modulation frequencies, such as those used in television, are necessary the value of R must be reduced to keep the RC time constant low enough to follow the envelope. Consequently the efficiency is reduced.

The diode detector can handle large signals without overloading, and it can provide automatic-volume-control voltage without extra tubes or special circuits. However, it has the disadvantage of drawing power from the input tuned circuit because the diode and its load form a low-impedance shunt across the circuit.

Consequently, the circuit Q, the sensitivity, and the selectivity are reduced. The interelectrode capacitance of the diode detector limits its usefulness at high carrier frequencies, and the bend in the lower portion of current-voltage characteristic indicates that it distorts on weak signals. Therefore considerable amplification is needed before detection.

Q5. In a superheterodyne receiver, why must the local oscillator tuning circuit track with the station selector tuning circuits?

38-14. Grid-Leak Detector

The GRID-LEAK-DETECTOR functions like a diode detector combined with a triode amplifier. It is convenient to consider detection and amplification as two separate functions. In Figure 38-16, A, the grid functions as the diode plate. The values of C_d and R_d must be so chosen that C_d charges during the positive peaks of the incoming signal and discharges during the negative peaks. The time constant of R_dC_d should be long with respect to the RF cycle and short with respect to the AF cycle.

An approximate analysis of the waveforms existing in the diode (grid) circuit is shown in Figure 38-16, B. Part 1 shows the input waveform which is also the waveform in the input tuned circuit. Because RF current i_g flows in only one direction in the grid circuit, part 2 shows a rectified current waveform in this circuit. Part 3 shows the waveform developed across C_d. This audio waveform is produced in the same way as the audio waveform in the diode detector. However, the waveform shown in part 3 is not the output voltage. In the grid-leak detector the waveform produced across C_d is combined in series with the RF waveform in the tuned circuit to produce the grid-to-cathode waveform shown in part 4 .

An approximate analysis of the waveforms existing in the triode plate circuit is shown in Figure 38-16, C. Part 5 is the plate-current waveform, and part 6 is the plate-voltage waveform.

Capacitor C discharges on the positive half cycles of grid input voltage (points 1, 3, 5, 7, 9, 11, 13, and so forth). The discharge path is clockwise through the circuit including the tube and capacitor. The time constant of the discharge path is the product of the effective tube resistance and the capacitance of capac-

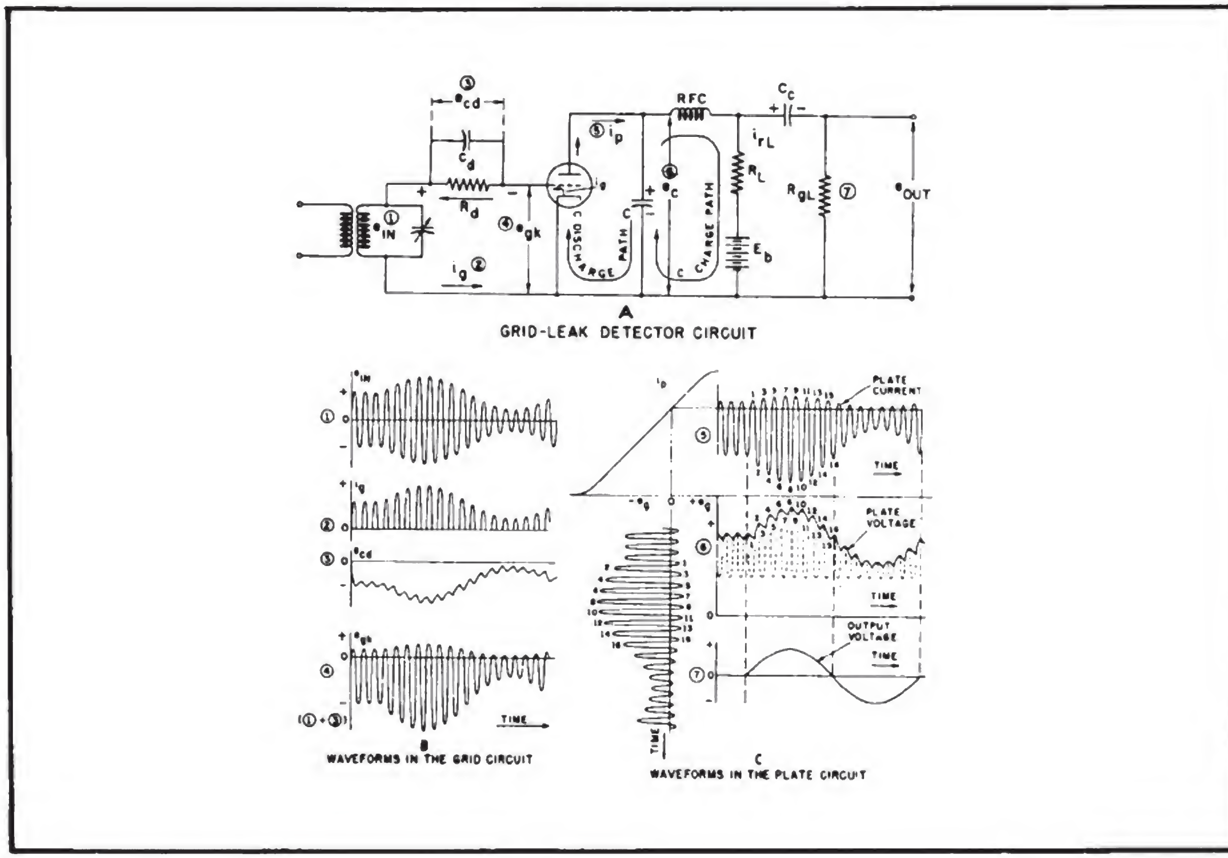

Figure 38-16 - Grid-leak detector and waveforms.

itor C, and this time constant is short because the effective tube resistance is low. The increase in plate current is supplied by the capacitor rather than the B supply, thus preventing any further increase in current through the RF choke and plate load resistor R_L. Therefore, any further change in plate and capacitor voltage is limited.

Capacitor C charges up as plate voltage rises on the negative half cycles of RF grid input voltage (Figure 38-16, C, points 2, 4, 6, 8, 10, 12, 14, and so forth). The charging path is clockwise through the circuit containing the capacitor, RF choke, load resistor R_L, and the B supply. The rise in plate voltage is limited by the capacitor charging current which flows through the RF choke and through R_L. The plate current decrease is approximately equal to the capacitor charging current; thus the total current through the RF choke and R_L remains nearly constant, and the plate and capacitor voltage rise is checked.

Positive grid swings cause sufficient grid current flow to produce grid-leak bias. Low plate voltage limits the plate current on no signal in the absence of grid bias. Thus, the amplitude of the input signal is limited, since with low plate voltage the cutoff bias is low, and that portion of the input signal that drives the grid voltage below cutoff is lost. The waveform of the voltage across capacitor C is shown by the solid line in part 6 of Figure 38-16, C. The plate voltage ripple is removed by the RF choke (RFC). Part 7 shows the output-voltage waveform. This waveform is the difference between the voltage at the junction of R_L and RFC with respect to the negative terminal of E_b and the voltage across coupling capacitor C_c, which for most practical purposes is a pure dc voltage.

Because the operation of the grid-leak detector depends on a certain amount of grid-current flow, a loading effect is produced which lowers the selectivity of the input circuit. However, the sensitivity of the grid-leak detector is moderately high on low-amplitude signals.

38-15. Plate Detector

In a grid-leak-detector the incoming RF signal is detected in the grid circuit and the

A5. The local oscillator tuning circuit must track with the station selector tuning circuits so that the oscillator will maintain a frequency which is 455kc above the station frequency over the entire tuning range.

---

resultant AF signal is amplified in the plate circuit. In a PLATE DETECTOR, the RF signal is first amplified in the plate circuit, and then it is detected in the same circuit.

A plate detector circuit is shown in Figure 38-17, A. The cathode bias resistor, R_1, is chosen so that the grid bias is approximately at cutoff during the time that an input signal of proper strength is applied. Plate current then flows only on the positive swings of grid voltage, during which time average plate current increases. The peak value of the ac input signal is limited to slightly less than the cutoff bias to prevent driving the grid voltage positive on the positive half cycles of the input signal. Thus, no grid current flows at any time in the input cycle, and the detector does not load the input tuned circuit, LC_1.

Cathode bypass capacitor C_2 is large enough to hold the voltage across R_1 steady at the lowest audio frequency to be detected in the plate circuit. C_3 is the demodulation capacitor across which the AF component is developed. R_2 is the plate load resistor. The RF choke blocks the RF component from the output. R_2C_3 has a long time constant with respect to the time for one RF cycle so that C_3 resists any voltage change which occurs at the RF rate. R_2C_3 has a short time constant with respect to the time for one AF cycle so that the capacitor is capable of charging and discharging at the audio rate.

The action of the plate detector may be demonstrated by the use of the i_p-e_g curve in Figure 38-17, B. On the positive half cycle of RF input signal (point 1) the plate voltage falls below the B supply because of the increased drop across R_2 and the RF choke. Capacitor C_3 discharges. The discharge current flows clockwise through the circuit including the tube and C_3. Plate current is supplied by C_3 rather than the B supply. The drop across R_2 and the RF choke is limited, and the decrease in plate voltage is slight.

On the negative half cycle of RF input signal (point 2) plate current is cut off and plate voltage rises. Capacitor C_3 charges. The charging current flows clockwise around the circuit including the RF choke, R_2, and the B supply. The drop across R_2 and the RF choke contributed by the charging current of C_3 checks the rise in plate voltage.

Thus, C_3 resists voltage change at the RF

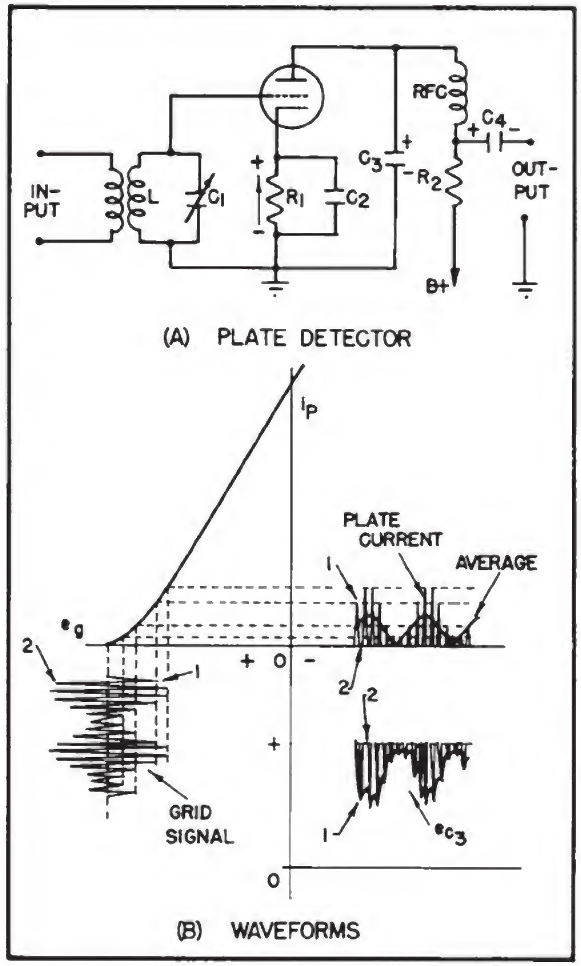

Figure 38-17 - Plate detector and waveforms.

rate. Because C_3R_2 has a short time constant with respect to the lowest AF signal, the voltage across C_3 varies at the AF rate.

The plate detector has excellent selectivity. Its sensitivity (ratio of AF output to RF input) is also greater than that of the diode detector. However, it is inferior to the diode detector in that it is unable to handle strong signals without overloading. Another disadvantage is that the operating bias will vary with the strength of the incoming signal and thus cause distortion unless a means is provided to maintain the signal input at a constant level. Thus, AVC or manual RF gain control circuits usually precede the detector.

38-16. Second Detector

Most superheterodyne receivers employ a diode as the second detector. This type of detector is practical because of the high gain as well as the high selectivity of the IF stages.

The diode detector has good linearity and can handle large signals without overloading. For reasons of space and economy, the diode detector and first audio amplifier are often included in the same envelope in modern superheterodyne receivers.

A simple diode detector is shown in Figure 38-18. The rectified voltage appears across R_1, which also serves as the volume-control potentiometer. Capacitor C_2 bypasses the RF component to ground, and C_3 couples the output of the detector to the first audio amplifier stage. The tuned circuit L_2C_1 is the secondary of the last IF transformer.

The time constant of R_1C_2 is long compared to the time for one IF cycle but short compared to the time for one AF cycle. If the intermediate frequency is 456 kc the time for one IF cycle in microseconds is

$$\frac{1}{0.456} = 2.19\,\mu s.$$

If R_1 is 250 k-ohms and C_2 is 100 pf the time constant in microseconds is

$$0.25 \times 100 = 25\,\mu s.$$

The demodulation capacitor, C_2, discharges through R_1 in one-half the time for one AF cycle $\frac{1}{2f}$. The time required to discharge C_2, is $5R_1C_2$ seconds. Thus,

$$\frac{1}{2f} = 5R_1C_2$$

$$f = \frac{1}{10R_1C_2}$$

$$= \frac{1}{10 \times 0.250 \times 10^6 \times 100 \times 10^{-12}}$$

$$= \frac{10^3}{0.250} = 4,000 \text{ cps}$$

Thus, the highest audio frequency which C_2 is capable of following without distortion is, in this example, 4,000 cps. In order to increase

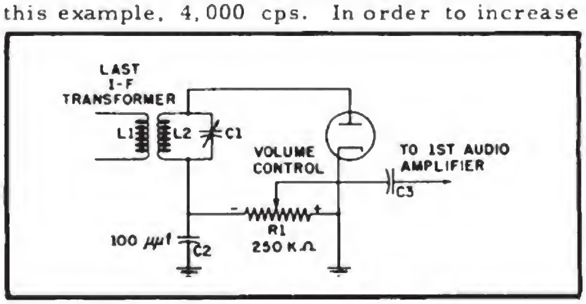

Figure 38-18 - Diode detector.

the response of the diode detector the time constant of R_1C_2 is reduced, for example, by decreasing R_1 to 100 k-ohms. The highest audio frequency now becomes

$$f = \frac{1}{10R_1C_2}$$

$$= \frac{1}{10 \times 0.100 \times 10^6 \times 100 \times 10^{-12}}$$

$$= \frac{10^6}{10^2} = 10,000 \text{ cps.}$$

Demodulation capacitor C_2 cannot discharge rapidly enough to follow modulation frequencies higher than 10,000 cps (in this case), and clipping results with all higher audio frequencies.

38-17. Circuit of a Superheterodyne Receiver

The complete circuit of a superheterodyne receiver is shown in Figure 38-19. In this circuit one RF amplifier (preselector) stage is used. Tube V_2, a pentagrid converter, serves both as the mixer tube and oscillator tube. Three tuning capacitors (one each in the preselector, mixer, and oscillator stages) are ganged on a common shaft to assure proper tracking. Trimmers are connected in parallel with each tuning capacitor to permit alignment. The oscillator tuning capacitor is smaller than the tuning capacitor in the preselector or the converter stages. The oscillator operates above the station frequency and tracks closely at three points on the dial—(1) low end, (2) middle, and (3) high end. The oscillator tuning capacitor split-rotor plates allow closer adjustment for tracking at the low end and at the middle of the band. Shunt trimmer capacity adjustments on the oscillator tuning capacitor provide close tracking of the oscillator at the high end of the band.

Tube V_3 is the IF amplifier with input and output IF transformers tuned to the receiver intermediate frequency.

Tube V_4 serves as the second detector and first audio amplifier. Conventional automatic volume control is tapped off at the end of the volume control potentiometer farthest from ground. Plate and screen potentials are obtained from the B supply through the corresponding voltage dropping resistors. The power supply is a conventional full-wave rectifier.

Q6. What limits the audio high frequency response of a diode detector?

Q7. What is the main disadvantage of the plate detector?

A6. The time constant of the RC filter network, which develops the audio signal, determines the upper audio frequency limit of the diode detector.

A7. The main disadvantage of the plate detector lies in its tendency to overload with a strong signal input.

The final stage in the receiver contains a power amplifier, it may be single ended or push-pull. Figure 38-19 shows a complete power amplifier stage containing V5, associated resistors and capacitors, a transformer, and a speaker. The stage operates as a class A amplifier in this case since it is single ended. A more complete coverage of audio power amplifier stages appears in Chapter 22. If push-pull is used, the stage may be operated class A, AB or B. In any case, the energy is transformer coupled to the load (speaker in Figure 38-19) for better impedance matching.

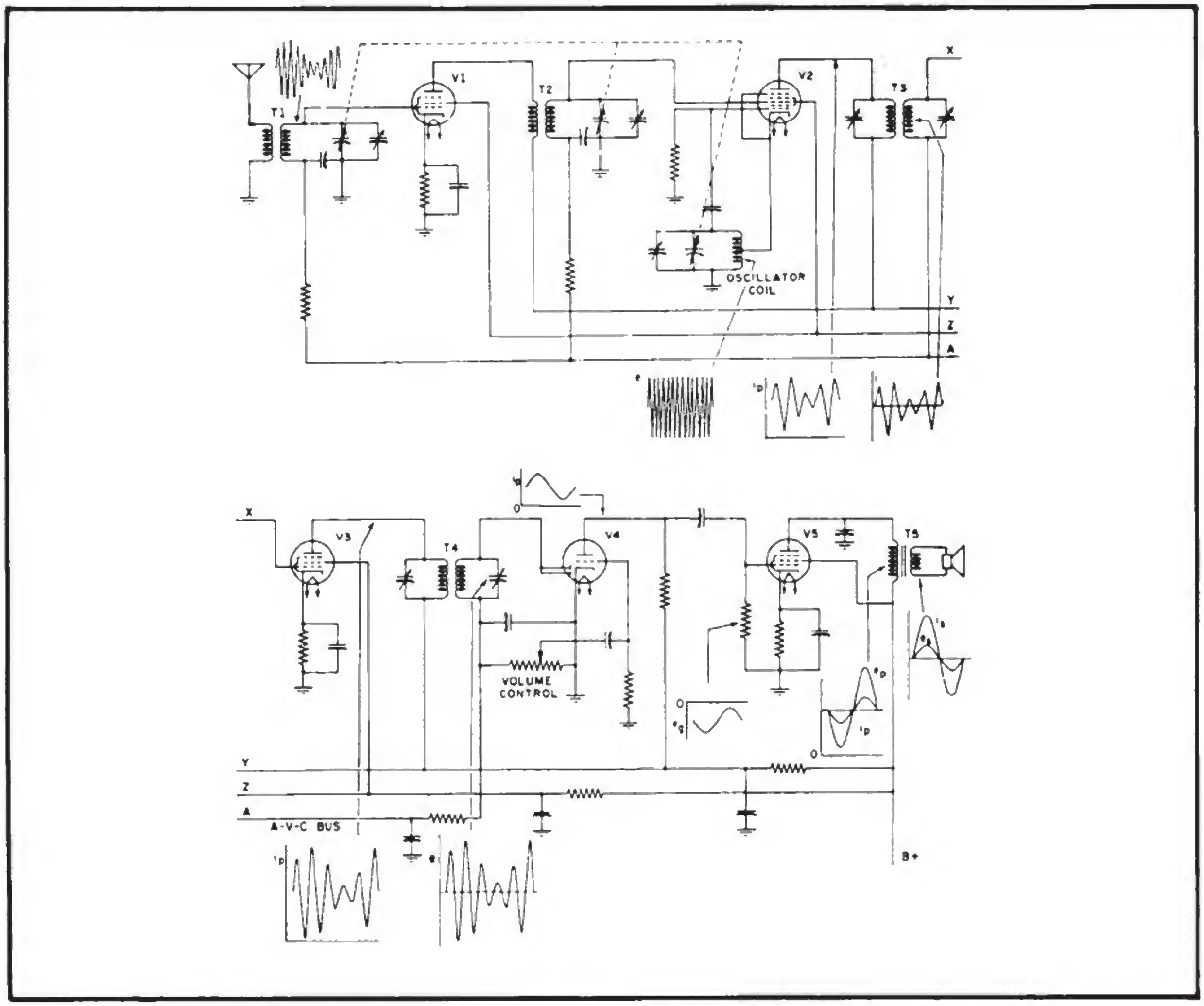

Figure 38-19 - Circuit diagram of a superheterodyne receiver.

EXERCISE 38

1. Which stage of the superheterodyne receiver establishes the selectivity and sensitivity of the receiver?
2. What is the function of the second detector?
3. What is the function of the decoupling circuits?
4. How does a tone control operate?
5. What causes motorboating in a receiver?
6. Draw the circuit for, and explain the operation of the superheterodyne receiver.
7. What is the difference between the TRF and the superheterodyne receiver?
8. What is an image frequency?
9. What is the function of a variable mu tube? How does it differ from a conventional pentode?
10. What is the difference between the first and second detector?
11. Describe the heterodyning principle.
12. Why must a demodulator be a nonlinear device?
13. Describe the operation of the grid leak detector.
14. What is the difference between the grid detector and the plate detector? What type of operation is used in the plate detector? How is the class of operation obtained?
15. What is conversion gain?

CHAPTER 39

RECEIVER CONTROL CIRCUITS

This chapter will deal with circuits which provide control of some receiver functions. These circuits can provide control, both manual and automatic, of such receiver functions as gain, local oscillator frequency, and tone of the audio output. The use of control circuits is not limited to radio receivers only. Many other forms of electronic equipments, such as Radar, Sonar, Direction Finders, and Navigational Aids, will use control circuitry to maintain optimum performance.

39-1. Manual Gain Control

Chapter 38 illustrated that high sensitivity is one of the parameters of a good receiver. In some cases, high sensitivity can be a liability. For example, the signal received from a nearby station can be strong enough to overload the RF sections of the receiver. This can cause the audio output to become distorted to the point of complete loss of intelligibility. To overcome this problem, gain control of the RF section is preferred rather than permanently decreasing receiver sensitivity. By using gain control, maximum sensitivity is realized, weak input signals are provided with maximum amplification yet when a strong input signal is received, the RF gain may be reduced to prevent overloading.

A typical manual gain control circuit is illustrated in Figure 39-1. RF amplifier V_1 is a variable Mu pentode. C_1 acts as the cathode by-pass capacitor. R_1 and R_2 develop cathode bias for V_1.

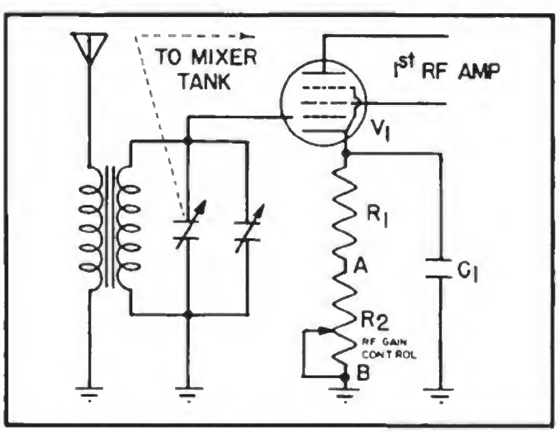

Figure 39-1 - Manual gain control.

The characteristics of variable Mu tubes were discussed in Chapter 38. It was illustrated that the gain of a remote cut-off pentode could be varied by changing the bias. Gain control R_2 (Figure 39-1) is nothing more than a manual bias control. When the wiper arm of R_2 is in position A, minimum bias is developed from cathode to ground, and the gain of V_1 is maximum.

When the wiper arm of R_2 is in position B, maximum bias is developed from cathode to ground, and the gain of V_1 is minimum. Fixed resistor R_1 is included to maintain a minimum value of bias on V_1 when the gain control is set at maximum and R_2 is effectively shorted out.

The disadvantage of this method is obvious. If signal strength keeps varying, the gain control must constantly be adjusted. An automatic gain control circuit overcomes this disadvantage.

39-2. Automatic Gain Control (AGC)

Under ideal conditions, once the manual volume or gain control has been set, the output signal should remain at the same level even if the input signals vary in intensity. The development of variable-mu tubes makes it possible to devise a practical avc or agc circuit, since the amplification of the tube may be controlled by varying the grid bias voltage. All that is needed is a source of bias voltage that varies with the signal strength. If this voltage is applied as bias to the grids of the variable-mu RF amplifier stages, the grids will become more negative as the signal becomes stronger. The amplification will thus be reduced, and the output of the receiver will tend to remain at a constant level. Unless the selectivity of the IF stages is good, strong adjacent channel signals will reduce receiver gain when a weak signal is tuned in. When no interference is present, avc holds the audio output constant as the input signal amplitude varies over a wide range.

The LOAD RESISTOR of a diode detector is an excellent source of this voltage, since the rectified signal voltage will increase and decrease with the signal strength. A filter is used to remove the AF component of the signal

and at the same time to prevent the avc circuit from shorting the audio output. Only the slower variations due to fading or change of position of the receiving antenna, and so forth, will then affect the gain of the RF amplifier stages because the avc circuit cannot compensate for very fast or extreme variations.

Figure 39-2 shows how the avc voltage is obtained. The avc voltage is tapped off at the negative end of the diode load resistor, R_2 (Figure 39-2, A), which is also the manual volume control. The AF component is removed by the filter circuit that is composed of C_2 and R_1. One or more of the RF amplifiers may be controlled by the voltage thus obtained. A customary value for R_1 is 2 megohms and for C_2 is 0.05 μf.

Figure 39-2, B, shows an avc circuit used with a duodiode triode in a conventional diode detector circuit. The two plates of the diode are connected together to form a half-wave rectifier in the RF portion of the circuit. The output of the diode detector is fed to the grid of the triode section which acts as a class A voltage amplifier.

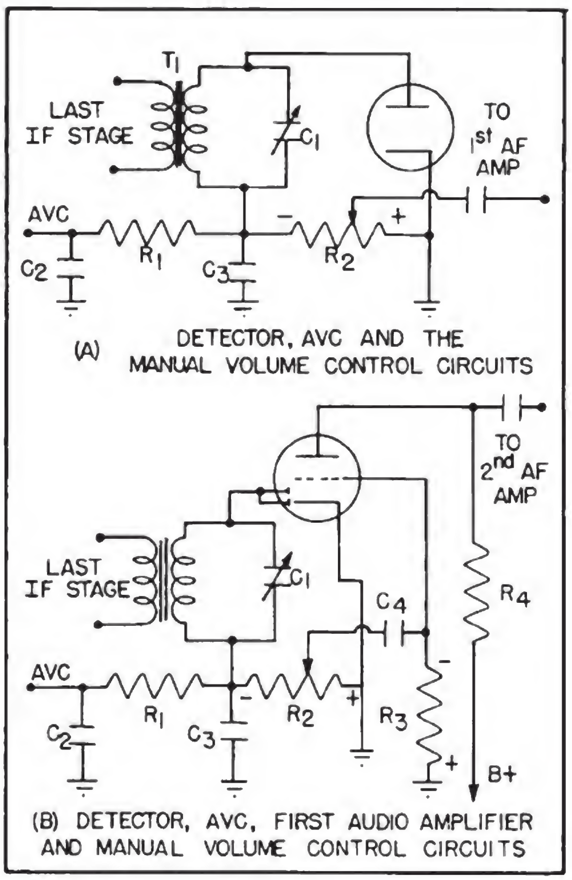

Figure 39-2 - Manual and avc circuits.

Low voltage bias is obtained by utilizing contact potential developed across R_3 resulting from the dissimilar elements in the grid and cathode.

A disadvantage of ordinary automatic volume control (avc) is that even the weakest signals produce some avc bias which could result in no usable receiver output or a reduction in overall receiver sensitivity. This reduction occurs because a slight increase in the avc will cause a decrease in RF amplifier gain, thereby, attenuating very weak signals.

39-3. Delayed Automatic Gain Control

The disadvantage of automatic gain control, that of attenuating even the very weak signals, is overcome by the use of delayed automatic gain control, as shown in Figure 39-3. In this circuit the avc diode, plate 2, is separated from the detector diode, plate 1, and both are housed in the same envelope with a triode amplifier.

In this example a bias of 5 volts on the delayed avc diode, plate 2, prevents it from conducting until the signal exceeds 5 volts. The signal across the secondary of the IF transformer is coupled to diode, plate 2, by capacitor C_1. Until the signal exceeds 5 volts no charge is acquired by the avc capacitor, C_3; no additional bias is applied to the grids of the IF amplifier, preselector, or converter tubes; and their gain is maximum on weak signals. The 5 volt bias applied to the delayed avc diode, plate 2, is developed across cathode resistor R_4 by the current flowing through the triode section of the tube. The triode section serves as a class A voltage amplifier driven by the audio voltage developed across diode load resistor R_2

When the signal across the secondary of the IF transformer exceeds the 5-volt bias value across R_4 the avc diode (plate 2) conducts on alternate half cycles and C_3 acquires a charge. The voltage developed across C_3 constitutes

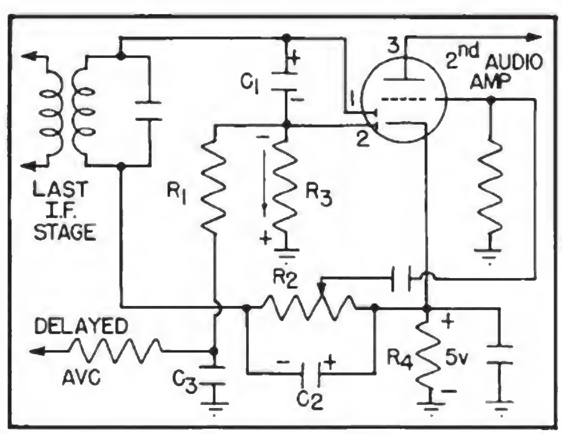

Figure 39-3 - Delayed avc circuits.

the delayed avc voltage. It is supplied to the grids of the various stages ahead of the second detector in series with the cathode bias developed by the individual tubes.

39-4. Beat-Frequency Oscillator

The beat-frequency oscillator (BFO) is necessary when CW signals are to be received because these signals are not modulated with an audio component. In superheterodyne receivers the incoming CW signal is converted to the intermediate frequency at the first detector as a single frequency signal with no side-band components. The IF signal is heterodyned (with a separate tunable oscillator known as the beat frequency oscillator) at the second detector to produce an AF output. In the circuit shown in Figure 39-4, the Hartley oscillator (BFO) is coupled to the plate of the second detector by capacitor C_3.

If the intermediate frequency is 455 kc and the BFO is tuned to 456 kc or 454 kc, the difference frequency of 1 kc is heard in the output. Generally the switch and capacitor tuning control are located on the front panel of the receiver.

The BFO should be shielded to prevent its own output from being radiated and combined with desired signals ahead of the second detector. If avc voltage is to be used it should be obtained from a separate diode isolated from the second detector. One way is to couple the output of an IF amplifier stage ahead of the second detector to the avc diode. Otherwise, the output of the BFO would be rectified by the second detector and would develop an avc voltage even on no signal.

39-5. Silencer

A silencer is sometimes employed in the AF section of a receiver to disable the receiver when no signals are being received. One type of silencer circuit is shown in Figure 39-5.

The silencer, V_1, a diode-connected triode, connects the output of the first AF stage to the input of the second audio amplifier. Silencer amplifier V_2 serves as the control tube for the silencer. The plate voltage of V_1 is supplied

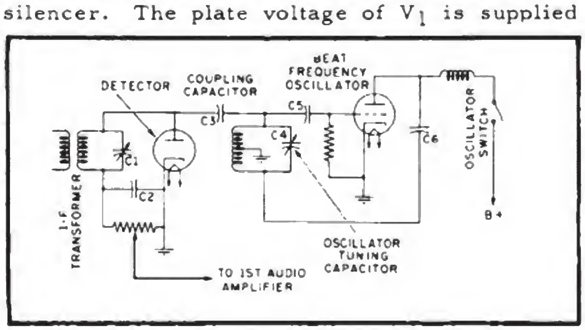

Figure 39-4 - Beat-frequency oscillator.

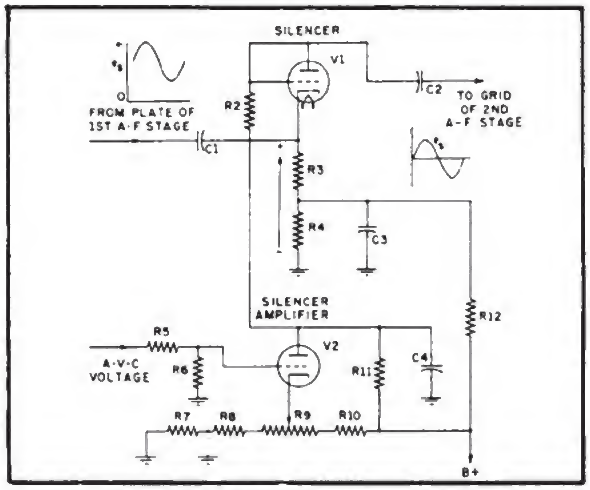

Figure 39-5 - Silencer circuit.

via R2 from the plate of V_2 (which is in turn supplied from the B+ supply via R_{11}) and is positive with respect to ground. The cathode voltage of V_1 is also positive with respect to ground, since it is connected to the B+ supply through a voltage divider made up of R_{12} and R4 With no input signal, R_9 is adjusted until V_2 draws enough plate current to reduce its plate voltage and that of V_1 to a value below the voltage on the cathode of V_1 Thus the silencer plate voltage is negative with respect to the cathode. Conduction ceases, and the silencer cuts off. The output is reduced to zero, and the receiver is mute.

The grid of V_2 is connected to the avc line. When a signal enters the receiver, the negative avc voltage is applied to the grid of V_2, thereby reducing the plate current and increasing the plate voltage of both V_2 and V_1. When the plate of V_1 becomes positive with respect to its cathode, the tube conducts and the signal is passed to the second AF amplifier.

Q1. What produces the delay action in a delayed automatic volume control circuit?

Q2. Why is avc voltage used to control the silencer circuit in a receiver?

39-6. Audio Tone Control

The tone of the sound reproduced in the audio section of the receiver is dependent on several factors. The frequency response of the audio amplifier will determine the degree of amplification provided to different frequencies in the sound spectrum. The size and quality of the loudspeaker will determine its response to various frequencies. The response of the human ear, which is the final judge of tonal quality, varies with the individual.

774-639 O - 65 - 12

A1. Delaying action, in an avc circuit, is produced by biasing the avc diode so that low amplitude signals will not produce avc voltage.

A2. The silencer circuit is controlled by the avc voltage, because avc voltage is produced only when an input signal is present. The silencer will then keep the audio section disabled until a signal is received.

Due to these variables, some form of tone control is usually employed in the audio amplifier. Treble tones are defined as the audio frequencies above approximately 3,000 cps. Bass tones are defined as the audio frequencies below approximately 300 cps. Although several methods of tone control can be used, only the attenuation method will be presented here. With this method, a decrease in the intensity of one tone can produce an apparent increase in the intensity of another tone. For example, two tones of 400 and 4,000 cycles are produced by a speaker with the same intensity. The intensity of the 4,000 cycle tone is reduced. The 400 cycle tone will now appear to be louder, although its intensity has not actually changed. Bass emphasis can then be accomplished by the attenuation of treble tones.

The simplest type of tone control is illustrated in Figure 39-6. Fixed capacitor C_1 by-passes the primary winding of the output transformer, effectively shunting the higher frequencies to ground. The size of C_1 will determine the lowest frequency to be affected. Values of from 0.001 to 0.05MFD are usually used. C_1 improves bass response by de-emphasizing the treble tones. This circuit is often used to improve the output of a small speaker with poor treble response.

A continuously variable tone control is illustrated in Figure 39-7. Tone control R_1 and by-pass capacitor C_1 act as a variable RC filter. With the wiper arm of R_1 in the bass position, C_1 by-passes the higher frequencies to ground, providing better bass response. When

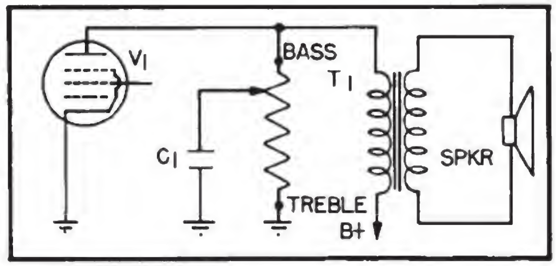

Figure 39-7 - Variable tone control using an RC filter.

the wiper arm of R_1 is in the treble position, the resistance of R_1 is placed in series with C_1, reducing the shunting effect of C_1 to high frequencies, and improving the treble response. This method has the advantage of providing smooth, continuous, tone control at all points between maximum bass and maximum treble response.

A switch type, variable tone control is illustrated in Figure 39-3. With this method, a three position switch is used to provide three fixed degrees of tone control. When tone control S_1 is in the bass position, capacitor C_1 by-passes the high frequencies and provides bass emphasis. With S_1 in the normal position, C_2 is acting as the by-pass, and a moderate amount of high frequency attenuation is accomplished. This position provides a relatively balanced bass and treble response. When S_1 is in the treble position, C_3 acts as the by-pass, providing minimum high frequency attenuation, and maximum treble emphasis. Note the relationship between the value of capacity used in each position and its effect on output tone.

39-7. Crystal Filter

A quartz crystal, used as a selective filter in the IF section of a communications receiver, is one of the most effective methods of achieving maximum selectivity. It is especially useful when the channel is crowded and considerable noise (both external and internal) is present. The crystal acts as a high Q tuned circuit, which is many times more selective than tuned

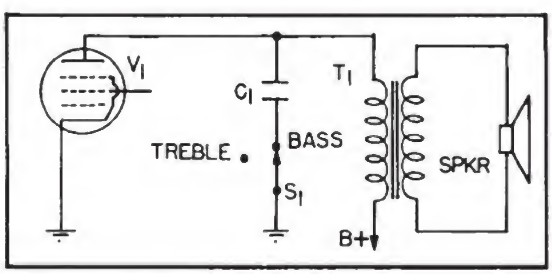

Figure 39-6 - Fixed capacitor tone control.

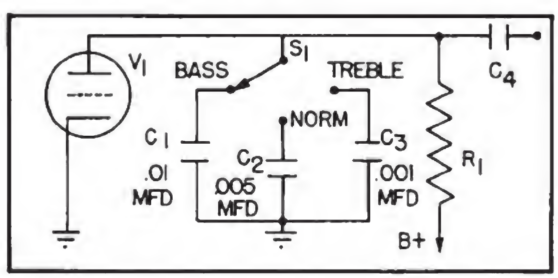

Figure 39-8 - Switch tone control.

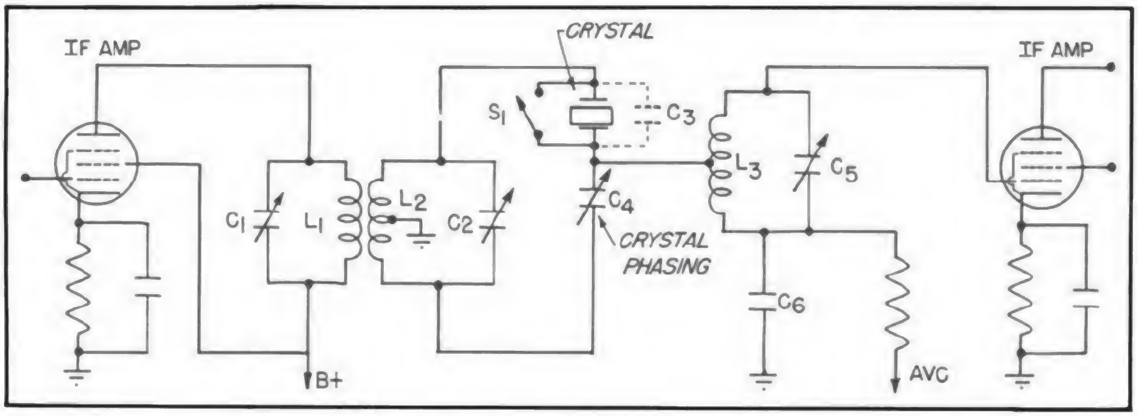

Figure 39-9 - Crystal filter used in the IF section of a superheterodyne receiver.

circuits consisting of inductors and capacitors. The crystal dimensions are so chosen that the crystal will be in resonance at the desired intermediate frequency.

One of the simplest of a number of possible circuit arrangements is shown in Figure 39-9. The crystal is in one arm of a bridge circuit. The secondary of the input transformer is balanced to ground through the center-tap connection. The phasing capacitor, C_4, is in another arm of the bridge circuit. The crystal acts as a high Q series resonant circuit and allows signals within the immediate vicinity of resonance to pass through the crystal to the output coil, L_3. The desired signal appears between the center tap of L_3 and ground.

The capacity between the crystal holder plates may bypass unwanted signals around the crystal. Therefore, some method must be provided to balance out this capacitance.

In this circuit, balancing is accomplished by taking a voltage 180° out of phase with the instantaneous voltage across the crystal and applying it via C_4 in such a way as to neutralize the undesired signal voltage. The balanced input circuit in this case is obtained by the use of a center-tapped inductor. The tap on L_3 permits the proper impedance match.

39-8. Automatic Frequency Control

Automatic Frequency Control (AFC) is used to maintain a constant frequency separation between the received signal and the local oscillator signal, regardless of drift in either the local oscillator frequency or the received carrier frequency.

The block diagram of an AFC system is illustrated in Figure 39-10. A change in the frequency of the LO will produce a resultant shift in the IF which will be detected by the discriminator. The discriminator will produce a dc voltage which corresponds to the shift in frequency. The polarity of this dc voltage will be determined by the direction of the frequency shift, while the amplitude of the dc voltage will be determined by the magnitude of the frequency shift. This dc voltage is applied to the control grid of a reactance tube, which in turn, controls the frequency of the oscillator tank circuit. The reactance tube acts as a variable capacitor in parallel with the oscillator tank, which will correct frequency drift by changing tank capacity.

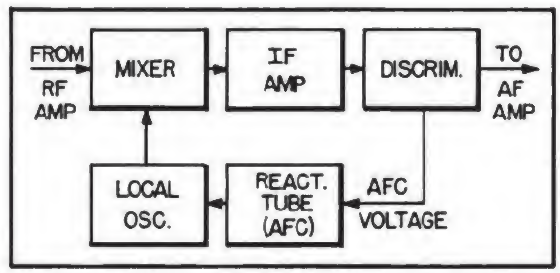

Figure 39-10 - AFC system - block diagram.

The theory of operation of a reactance-tube circuit may be explained with the aid of Figure 39-11. The reactance tube, V_1 (Figure 39-11, A), is effectively in shunt with the oscillator tank, LC, and the phase shift circuit, R_gC_1. The capacitive reactance of the capacitor is large compared with the resistance of the resistor; and the current, i, in this circuit leads the voltage, e_p, across the circuit by approximately 90 degrees. The voltage, e_p, is the alternating component of the plate to ground voltage appearing simultaneously across the reactance tube, the phase-shift circuit, and the oscillator tank.

The coupling capacitor, C_2, has relatively low capacitive reactance to the ac component of

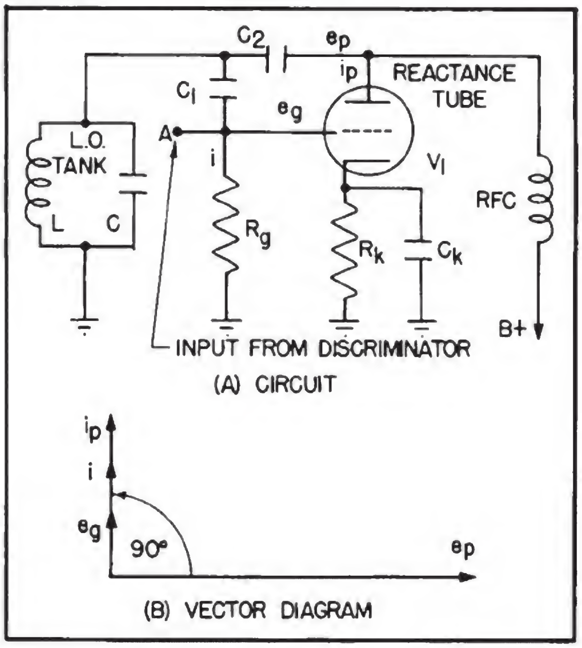

Figure 39-11 - Reactance tube AFC.

current through it, and at the same time it blocks the dc plate voltage from the phase shift circuit and the tank. The reactance tube receives its ac grid input voltage, e_g, across R_g. This voltage is the IR drop across R_g and is in phase with plate current i_p. This relation is characteristic of amplifier tubes.

Because e_g is in phase with both i and i_p and e_g leads e_p by approximately 90°, both i and i_p lead e_p by approximately 90 degrees. These relations are shown in the vector diagram in Figure 39-11, B. Both i and i_p are supplied by the oscillator tank circuit, and because both are leading currents with respect to the tank voltage, e_p, they act like the current in tank capacitor C. Therefore, the EFFECT of these currents on the frequency of the tank is the same as though additional capacitance were connected in parallel with it.

Consider the effect of applying a dc voltage to point A in Figure 39-11. When the local oscillator is operating at the correct frequency, the dc output of the discriminator is zero, and consequently no dc voltage is applied to the grid (point A) of V_1. The plate current of V_1 is a succession of constant amplitude pulses, and the circuit is operating at the no-signal condition.

If the local oscillator frequency was to increase, a positive dc voltage would be produced by the discriminator. The amplitude of the dc voltage would be determined by the amount of frequency deviation. This positive dc voltage is applied to the grid of V_1, causing an increase in the amplitude of V_1 plate current. Since the reactance tube circuit appears as a capacitor in shunt with the LO tank, an increase in V_1 plate current causes this apparent capacity to seem larger. This increase in tank capacity would decrease the LO frequency until the input voltage to V_1, or discriminator output, was again equal to zero.

If a decrease in LO frequency occurred, a negative voltage would be produced by the discriminator. The amplitude of the dc would once again be determined by the degree of frequency deviation. This negative dc voltage, which is applied to the grid of V_1, would cause a decrease in the plate current of V_1. This decrease in plate current would decrease the apparent capacity in shunt with the LO tank, increasing the LO frequency.

Q3. What is the advantage of using a crystal filter in the IF section of a receiver rather than in the RF section?

Q4. What effect will the AFC circuit have on the LO when the F-M receiver is tuned from one station to another?

EXERCISE 39

1. Describe the operation of the manual gain control.
2. What is the disadvantage of the manual gain control?
3. Describe the operation of the automatic gain control.
4. What is the disadvantage in the use of the ordinary gain control?
5. Describe the operation of a delayed avc circuit.
6. What is a silencer circuit? What is its function?
7. What could be added to an ordinary audio amplifier to provide tone control?
8. Describe the operation of the reactance tube AFC circuit.
9. The crystal filter used in the IF section is cut to what frequency. Why is it used?
10. What is a beat frequency oscillator? For what function may it be used?

A3. A crystal used in the IF section of a receiver will provide filtering for all stations in the receiver tuning band. When used in the RF section, a separate crystal must be provided for each station frequency.

A4. The AFC circuit will have no effect because the IF remains constant over the entire tuning range.

CHAPTER 40

FREQUENCY DEMODULATION

The intelligence to be transmitted may be superimposed on the carrier in the form of changes in the frequency of the carrier. This type of modulation is called FREQUENCY MODULATION and has certain inherent advantages over conventional AM transmission, particularly when static free transmission is desired. However, in extensive tests conducted at the Naval Research Lab (NRL), it was found that for general Navy use amplitude modulation was in many ways more desirable than narrow band frequency modulation. Nevertheless, the Naval Communication System uses a limited number of FM transmitters and receivers. Aircraft altimeters use frequency modulation, as do some other radar and sonar equipments.

Intelligence may be conveyed by varying the frequency of a continuous radio wave of constant amplitude. The carrier frequency can be varied a small amount on either side of its average, or assigned, value by means of the AF modulating signal. The amount the carrier is varied depends on the amplitude of the modulating signal, and the frequency with which the carrier is varied depends on the frequency of the modulating signal. The amplitude of the RF carrier remains constant with or without modulation.

There are several systems of frequency modulation that fulfill these requirements. A MECHANICAL MODULATOR employing a capacitor microphone is the simplest system of frequency modulation. Two other systems of frequency modulation are REACTANCE-TUBE and ANGLE, or PHASE-ANGLE, MODULATION. The main difference between these two systems is that in reactance-tube modulation the RF wave is modulated at its source (the oscillator), while in phase modulation the RF wave is modulated in some stage following the oscillator. The results of each of these systems are the same, that is, the FM wave created by either system can be received by the same receiver.

40-1. Capacitor-Microphone System

The simplest form of frequency modulation is that of a capacitor microphone, which shunts the oscillator tank circuit, LC, as shown in Figure 40-1. The capacitor microphone is equivalent to an air dielectric capacitor, one plate of which forms the diaphragm of the microphone. Sound waves striking the diaphragm compress and release it, thus causing the capacitance to vary in accordance with the spacing between the plates. This type of transmitter is not practicable (among other reasons, the frequency change is very limited), but it is useful in explaining the principles of frequency modulation. The oscillator frequency depends on the inductance and capacitance of the tank circuit, LC, and therefore varies in accordance with the changing capacitance of the capacitor microphone.

If the sound waves vibrate the microphone diaphragm at a low frequency, the oscillator frequency is changed only a few times per second. If the sound frequency is higher, the oscillator frequency is changed more times per second. When the sound waves have low amplitude, the extent of the oscillator frequency change from the no signal, or resting, frequency is small. A loud AF signal changes the capacitance a greater amount and therefore deviates the oscillator frequency to a greater degree.

Thus, the frequency of the AF signal determines the number of times per second (RATE OF DEVIATION) that the oscillator's tank frequency changes. The amplitude of the AF signal determines the extent of the tank circuit frequency change (AMOUNT OF DEVIATION).

A graph representing the changing tank circuit frequency at an audio rate is shown in Figure 40-2.

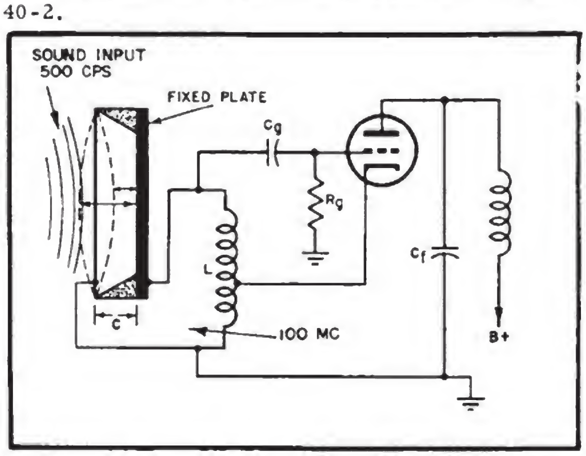

Figure 40-1 - FM transmitter modulated by a capacitor microphone.

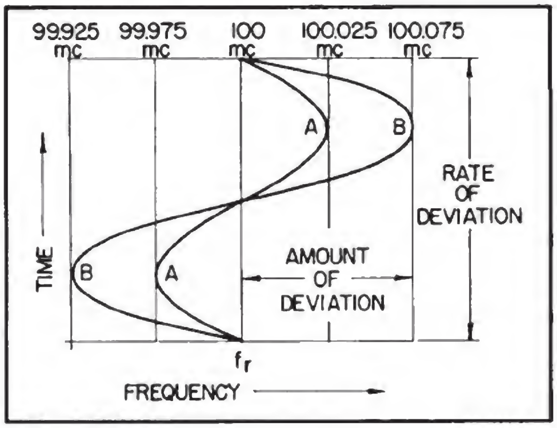

Figure 40-2 - Deviation changes in FM.

f_r represents the resting frequency of the tank circuit when there is no modulating signal present, and the diaphragm of the capacitor microphone is motionless (in this case f_r = 100 Mc). With a sinusoidal sound source of 500 cps in (A in Figure 40-2), the amount of deviation on either side of the resting frequency is 25 kc and the rate of deviation is 500 cps.

When the amplitude of the 500 cps sound source is increased to B (as shown in Figure 40-2) the microphone diaphragm vibrates farther on either side of its resting position but still crosses its resting position the same number of times per second. In this case the rate of deviation remains constant but the amount of deviation increases to 75 kc. The oscillator frequency now changes from 99.925 Mc to 100.075 Mc.

It is also possible to keep the amount of deviation constant but vary the rate of deviation. In a practical FM transmitter the two independent variables, rate and amount of deviation, are continually changing. This occurs because the amplitude (strength) and frequency of the modulating signal is continually changing.

Figure 40-3 shows the actual tank circuit output voltage compared (with respect to time) to the modulating signal as might be seen on an oscilloscope.

The horizontal axis represents a linear time base. The degree (how high or how low in frequency) of compression or rarefaction is dependent upon the amplitude of the audio signal. The number of compressions or rarefactions per second would be determined by the audio frequency. Thus it is seen how RF energy can carry intelligence in its as frequency changes.

Q1. In FM, upon what does the deviation frequency depend?

Q2. What characteristic of the modulating wave determines the number of times per second that the frequency deviates about the resting frequency?

40-2. Reactance Tube System

In Chapter 39, it was shown how a dc control voltage could change the frequency of an oscillator by the use of a reactance tube. A more practical method of obtaining FM is by use of a reactance tube, where the reactance tube is controlled by an audio signal instead of a dc control signal.

The reactance tube system of frequency modulation is shown in Figure 40-4. The reactance tube is an electron tube operated so that its reactance varies with the modulation signal and thereby varies the frequency of the oscillator stage.

In this circuit the reactance tube is connected in parallel with the oscillator tank and functions like a capacitor whose capacitance is varied in accordance with the audio signal, as in the capacitor microphone system of frequency modulation.

The frequency of the AF signal determines the number of times per second that the oscillator tank frequency changes. On the other hand, the amplitude of the AF signal determines the extent of the oscillator frequency change that is, the amount of deviation. The frequency of the oscillator is thus changed, and the resulting FM signal is passed through a frequency doubler to increase the carrier frequency and the deviation frequency. A power amplifier feeds the final signal to the antenna. The transmitter is kept within its assigned frequency limits by comparing the output of the transmitter with that of a standard crystal controlled oscillator, and feeding back a suitable correcting voltage from a frequency converter and discriminator (frequency detector) stage.

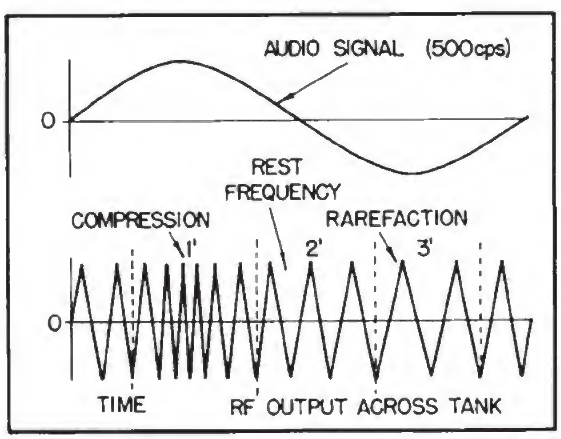

Figure 40-3 - FM waveform.

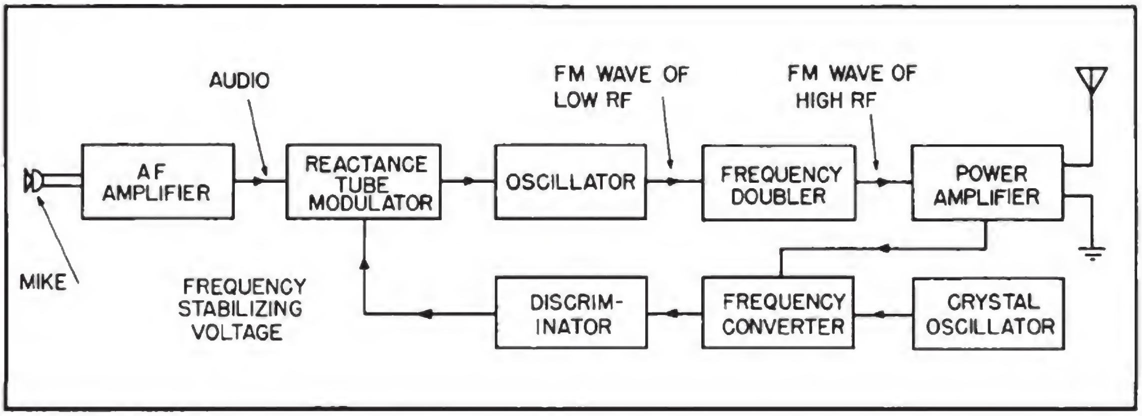

Figure 40-4 - Block diagram of a reactance-tube FM transmitter.

Q3. In the reactance-tube FM transmitter what stage is responsible for increasing the frequency range of the transmitter?

40-3. FM Sidebands

With the application of one audio frequency to an AM transmitter, two sidebands are formed. Their position in the frequency spectrum was above and below the carrier frequency by an amount equal to the audio frequency. During this process of amplitude modulation, carrier power remains constant but the relative amplitube (power) of the sidebands varies in accordance with the intelligence being transmitted. This additional power is developed in the modulator.

In FM, the application of a modulating signal causes the carrier frequency to vary about the resting frequency determined by the amplitude of the modulating signal. In the process of varying, sidebands are formed and theoretically an FM wave with single tone modulation has an infinite number of sideband pairs instead of just one pair, as in amplitude modulation. However, only a limited number of sidebands contain sufficient energy to be significant.

A significant sideband is defined as a sideband that has sufficient power to be of value toward the reception of the transmitted signal. Usually, if a sideband contains less than 1% of the total carrier power, it is considered insignificant.

Figure 40-5 represents the FM wave consisting of a carrier wave of frequency (fr) and associated sideband frequencies of fr±fm, fr±2fm, fr±3fm, and so forth, where fm is the modulating frequency and fr is the resting frequency. The lines to the right of center represent the upper sideband frequency components, and those to the left of center represent the lower sideband components. The lengths of the lines represent the energy levels of the various components. The sideband frequency components are spaced an amount equal to the modulating frequency. For example, if a 500 cps signal is applied to the microphone of an FM transmitter, sidebands will appear above and below the rest frequency, each sideband spaced 500 cps apart from the sideband next to it. The audio frequency fixes the separation of the sidebands.

Q4. What is the relationship between the sidebands produced in FM and the audio modulating frequency?

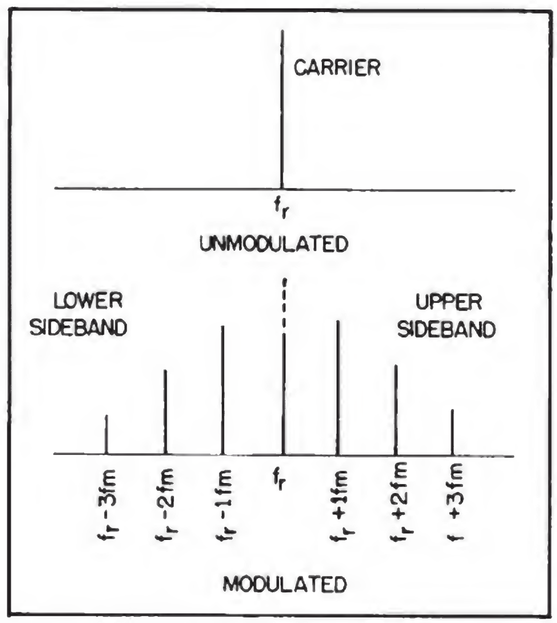

Figure 40-5 - An FM signal with its sidebands developed as a result of audio modulation.

A1. The distance between the sound source and the microphone and the intensity (strength) of the sound source.

A2. The frequency of the modulating signal.

A3. The frequency doubler stage.

A4. The sidebands are separated by the audio frequency.

40-4. Sideband Power

The FM wave consists of a center or resting frequency and a number of sideband pairs which, for a given audio frequency and amplitude, are constant. However, the resultant wave varies in frequency but is constant in amplitude. This resultant wave is the algebraic sum of the components which form it, and the carrier or resting frequency will vary in amplitude with the modulation. When the transmitted signal is unmodulated, there is a certain constant amount of power in the carrier signal. When modulation is applied, power is taken from the carrier and forced into the bands; therefore, the carrier amplitude or resting frequency component is reduced.

In the FM wave the carrier frequency or resting frequency component changes in amplitude with modulation applied, whereas in the AM wave the carrier component remains constant in amplitude with modulation applied. In FM transmission the power in the sidebands subtracts from the power in the carrier, whereas in AM transmission the power for the sidebands is supplied by the modulator and is not subtracted from the carrier. Since the carrier contains no intelligence, reducing its amplitude will increase efficiency (less power consumed).

40-5. Modulation Index

Whenever sidebands are formed, they are spaced by an amount equal to the frequency of the modulating signal. A modulating signal with a frequency of 5 kc results in sidebands spaced at 5 kc intervals, whereas, for a frequency of 10 kc, the sidebands are spaced 10 kc apart. As with amplitude modulation, the bandwidth for frequency modulation is determined by the number of sidebands associated with the carrier. If the amplitude of the two modulating signals are causing 8 significant sidebands on each side of the carrier, the 5 kc note would result in a bandwidth of 8x5 or 40 kc, on each side of the carrier, or a total of 2x40 or 80 kc. On the other hand, with the 10 kc note, the total frequency spectrum would extend 160 kc. The bandwidth thus depends upon (1) frequency of the modulating signal and (2) the total frequency deviation of the carrier (amplitude of the modulating signal). The ratio between the frequency deviation and the modulation frequency is termed the MODULATION INDEX. The modulation index number denotes the number of pairs and relative amplitude of sideband components produced during frequency modulation.

$$\text{Modulation Index (M)} = \frac{\text{deviation of FM carrier } (f_d)}{\text{frequency of modulating signal } (f_m)} \qquad (40\text{-}1)$$

A comparison between the modulation index to the number of significant sideband pairs and their relative magnitude are shown in the frequency spectrum graph illustrated in Figure 40-6. The modulation index is an instantaneous value because the strength and frequency of the modulating signal, in a practical transmitter, is continually changing.

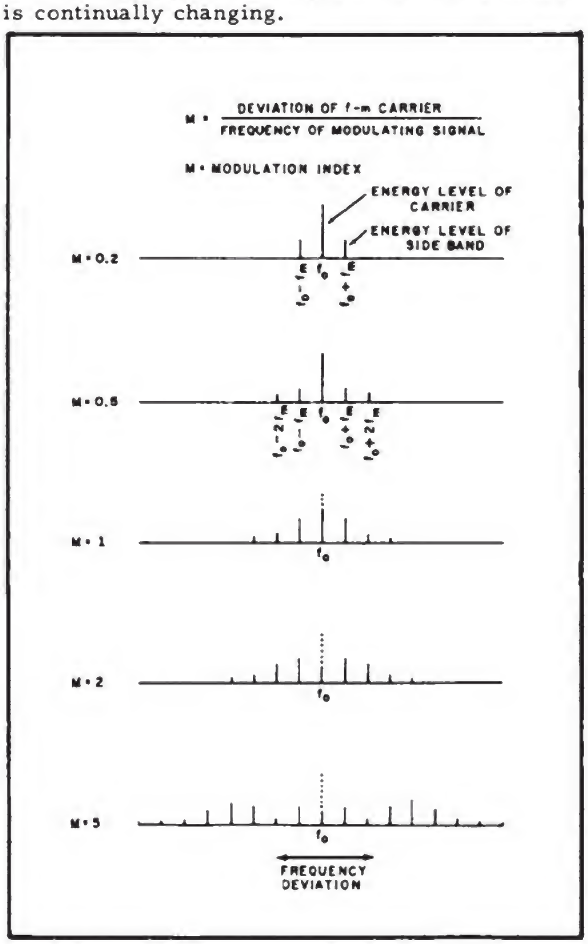

Figure 40-6 - Frequency and energy distribution for 5 values of modulation index of an FM wave.

This formula illustrates that, if the frequency of the modulating signal is held constant, the bandwidth will depend directly upon the amplitude of the modulating signal. On the other hand, if the amplitude is held constant, the lower audio frequencies will produce more sidebands. For commerical high fidelity broadcast transmission, 15 kc is the highest modulation frequency in use. The maximum frequency deviation of the carrier is limited by the Federal Communications Commission (FCC) to 75 kc on each side of the carrier frequency. The ratio between these two maximum limits is called the DEVIATION RATIO where:

$$\text{deviation ratio} = \frac{\text{maximum amount of deviation}}{\text{maximum frequency of modulating signal}} \qquad (40\text{-}2)$$

Although the FCC regulation limits the amount of deviation for a commercial FM broadcast transmitter to 75 kc, some significant sidebands may extend beyond this frequency. A guard band of 25 kc on each side of the allowable frequency swing of ± 75 kc is established to take care of most of the significant sidebands beyond the established limits.

The modulation index number by itself does not directly tell us anything. The calculation of the actual number and relative amplitude of the various sidebands requires the use of Bessel functions which are complex. Modified versions of Bessel function charts are given in Table 1 which lists the number of significant sidebands that would be obtained for some of the common values of the modulation index. Thus, with an index of 2, there are 4 significant sidebands (on either side of the carrier) formed, with an index of 5, the number of sidebands increases to 8. In general, the higher the modulation index the more significant sidebands.

When the modulation index becomes 0.4 or less only two sidebands are formed which is similar to AM when one audio signal is used. In all cases, as modulation index increases in value, the carrier power decreases.

Q5. What determines the effective bandwidth of an FM transmitter.

Q6. With an amount of deviation equal to 75 kc and an audio frequency of 15 kc, what is the approximate bandwidth of the transmitted signal?

Q7. What are guardbands and why are they required in FM transmission?

Modulation Index	Number of effective Sidebands	
	Above carrier	Below carrier
0.1	1	1
0.2	1	1
0.3	1	1
0.4	1	1
0.5	2	2
1.0	3	3
2.0	4	4
3.0	6	6
4.0	7	7
5.0	8	8
6.0	9	9
7.0	10	10
8.0	12	12
9.0	13	13
10.0	14	14
11.0	16	16
12.0	17	17
13.0	18	18
14.0	19	19
15.0	20	20

40-6. Degree of Modulation

To explain 100 percent modulation in an FM system, it is desirable to first review the same condition for an AM wave. As has been stated, 100 percent modulation (AM) exists when the amplitude of the envelope varies between zero and twice its normal unmodulated value. There is a corresponding increase in power of 50 percent. The amount of power increase depends upon the degree of modulation; and because the degree of modulation varies, the tubes cannot be operated at maximum efficiency continuously.

In frequency modulation, 100 percent modulation has a different meaning. The AF signal varies only the frequency of the oscillator. Therefore, the tubes operate at maximum efficiency continuously and the FM signal has a constant power input at the transmitting antenna regardless of the degree of modulation. A modulation of 100 percent simply means that the carrier is deviated in frequency by the full permissible amount. For example, an 88 Mc FM station has 100 percent modulation when its audio signal deviates the carrier 75 kc above and 75 kc below the 88 Mc value, when this value is assumed to be the maximum permissible frequency swing. For 50 percent modulation, the frequency would be deviated 37.5 kc above and below the resting frequency.

Q8. If the maximum permissible amount of deviation is 75 kc and the actual amount of deviation is 25 kc, what is the percent of modulation?

A5. The amount of deviation and the frequency of the modulating signal.

A6. Approximately 240 kc.

A7. Frequency space on either side of the deviation limits to prevent adjacent channel interference.

A8. 33.3 percent.

40-7. FM Receivers

The TRF and superheterodyne receivers that have been described in the preceding chapters are designed to receive RF signals that vary in amplitude according to the audio modulation at the transmitter. The amplitude of the RF signal is increased by one or more RF amplifier stages, and the modulation component is reproduced by the detector. Each of the tuned circuits preceding the detector is designed to pass only a relatively narrow band of frequencies containing the necessary upper and lower sideband frequencies associated with the amplitude modulated carrier.

FM receivers are supplied RF signals that vary in frequency according to the information being transmitted. The amount of the variation or deviation from the CENTER, or RESTING FREQUENCY at a given instant depends on the amplitude of the impressed audio signal. The frequency with which the variations from the center frequency occur depends on the frequency of the impressed audio signal. The function of the FM receiver is basically the same as that of the AM superheterodyne receiver, that is, the amplitude of the incoming RF signals is increased in the RF stages; then the frequency is reduced in the mixer stage to the intermediate frequency and amplified in the IF amplifier section. Finally, the amplitude is clipped in the limiter stage and the modulation component is reproduced by the detector, or DISCRIMINATOR as it is called in the FM receiver.

There are a few major differences between the FM and the AM receiver. The greatest difference is in the method of detection. Also the tuned circuits of the FM receiver have a wider bandpass and the last IF stage is especially adapted for limiting the amplitude of the incoming signal. However, in both systems the audio amplifiers and reproducers are similar.

A superheterodyne receiver designed for FM reception is shown in Figure 40-7.

The function of the FM antenna is to provide maximum signal voltage to the receiver input. FM antennas are cut to the required length in order to receive a signal of sufficient amplitude to drive the first RF amplifier.

If a single frequency is to be received, the antenna may be designed for maximum response at that frequency. If, however, a band of frequencies is to be received, the antenna length will represent a compromise. Usually the length is so chosen that it will be in resonance at the geometric center of the band. The GEOMETRIC CENTER or MEAN is equal to X_1X_2, where X_1 and X_2 are the wavelengths at the two ends of the band.

There are many types of FM antennas, but probably the simplest is the half-wave dipole. The length of the half-wave dipole, in feet, is

$$\frac{468}{f_1 \; f_2} \qquad (40\text{-}3)$$

where f_1 and f_2 are the frequencies in megacycles at the two ends of the band. Because the resistance at the center of the half-wave dipole is about 72 ohms, the transmission line connecting the antenna with the receiver should have a characteristic impedance of 72 ohms in order to operate as a nonresonant transmission line with no standing waves. The transmission line feeds the signal to the receiver via a matching transformer at the input to the preselector stage.

The RF amplifier, or preselector, performs

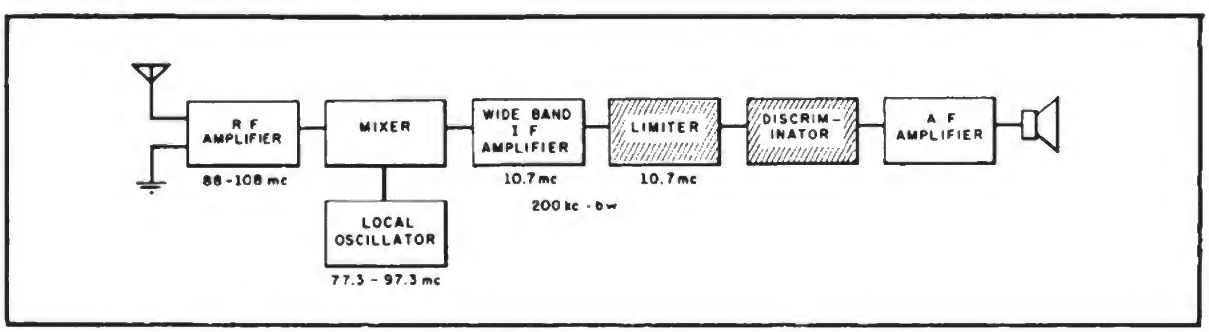

Figure 40-7 - Receiver block diagram.

essentially the same function in the FM receiver as it does in the AM receiver, that is, it increases the sensitivity of the receiver. Such an increase in sensitivity is often a practical necessity in fringe areas. However, the gain of the IF stages is much greater, perhaps 10 times that of the preselector, since the chief advantage of the superheterodyne lies in the uniformity of response and gain of the IF stages within the receiver band. The principle functions of the RF stage are to discriminate against undesired signals (images) and to increase the amplitude of weak signals so that the signal-to-noise ratio will be improved.

If the receiver is designed to receive both amplitude modulation and frequency modulation, a suitable band switching arrangement is necessary. Many combination receivers are designed to receive more than one AM band. Under such circumstances additional tuned circuits are needed. Thus, if two AM bands and one FM band are used and one RF stage is used ahead of the mixer, three tuned circuits are needed for each band to be covered. This circuit arrangement includes one for the RF stage, one for the mixer stage, and one for the oscillator stage for each of the three bands, or a total of nine tuned circuits. The FM tuned circuits have wider bandpass characteristics than do the AM tuned circuits, as shown in Figure 40-7.

Q9. What is the difference between an antenna used for a broadcast AM receiver and one used for a broadcast FM receiver.

Q10. Why does commerical FM have limited coverage?

40-8. Frequency Converter

The frequency converter employed in the FM receiver functions in much the same manner as the one employed in the AM superheterodyne receiver. However, additional problems are involved.

For example, at the FM broadcast frequencies employed in the commerical FM band the stability of the local oscillator becomes a major problem. There is a tendency for the local oscillator to become synchronized with the incoming signal and thus to lose the intermediate frequency output entirely. The tendency is more pronounced at FM frequencies because the station and oscillator are relatively closer together. Therefore, for maximum frequency stability, separate oscillator tubes are used. This results in increased space requirements and expense. Especially designed pentagrid converters that have reasonably good frequency stability, high conversion transconductance, and oscillator transconductance are employed in some less expensive commerical sets.

Even in a normal well designed FM receiver such factors as the change in internal capacitance of the oscillator tube (or oscillator section of a tube) and the expansion of coil windings and capacitor plates during warmup may cause the local oscillator frequency, and consequently the intermediate frequency, to drift an appreciable amount. A relatively small shift in oscillator frequency may shift the IF signal beyond the range of the IF stages with a consequent loss in output signal.

Various methods are used to combat oscillator drift. For example, the second harmonic of the local oscillator frequency is sometimes used for mixing. In this instance the local oscillator may be operated at a lower fundamental frequency, where the stability is improved. Another method is to use capacitors having a negative temperature coefficient. Capacitors normally have a positive temperature coefficient (increase in temperature results in increase in capacitance). Capacitors with negative temperature can also be constructed by special design. These are connected in shunt with capacitors having a positive temperature coefficient to counteract the change in capacitance when the temperature of the oscillator stage varies. Proper voltage regulation as well as the choice of oscillator tubes having low internal capacitances, will also increase the stability of the local oscillator.

Frequency stability of the local oscillator, in the standard FM band, makes it advantageous to operate the local oscillator at a frequency below that of the incoming signal. (See Figure 40-7.

However, if the local oscillator is operated above the frequency of the incoming signal it is not so likely to interfere with television receivers in the same vicinity that are operating on the lower video channels. Therefore, some commerical FM receivers have local oscillators operating above the incoming signal.

Q11. Why is a mixer used in preference to a converter stage in an FM receiver?

Q12. What methods are used to compensate for local oscillator drift?

40-9. IF Amplifier

The IF amplifier in an FM receiver is usually tuned to a center frequency of from 8 to 10 megacycles. It generally employs double tuned transformers having equal primary and secondary inductances. The bandpass is from 150 to 200 kc. The last one or two IF stages function as a limiter.

A9. Due to the high frequencies involved the FM antenna is usually cut to a certain length.

A10. Because the VHF band (88-108 Mc) is subject to line of sight communication.

A11. Less change for synchronization between the oscillator section and the incoming signal to occur.

A12. Automatic frequency control circuits, proper voltage regulation, oscillator tubes having low internal capacitance.

The gain of each wide band IF stage is considerably less than that of the narrow band AM type of IF amplifier. Therefore, an FM receiver employs more IF stages than a corresponding AM receiver.

A low value of intermediate frequency is undesirable because local oscillator drift might force the set to operate outside the IF range. Also, it would be pointless to have the intermediate frequency lower than the total frequency deviation (bandwidth) of any one FM station.

In the choice of the optimum IF value such factors as image response, response to signals at the same frequency as the intermediate frequency, response to beat signals produced by two stations separated in frequency by the IF value, and response to harmonic frequencies must be considered.

Two stations separated in frequency by the IF value will, if sufficiently powerful, produce a beat frequency that will pass through the receiver. This type of interference may be eliminated if the intermediate frequency chosen is greater than the entire FM bandwidth. It may be minimized by adequate discrimination in the preselector stage.

Harmonics of the local oscillator may combine with harmonics produced when a strong incoming signal overloads the input stage to produce the intermediate frequency.

Interfering signals may develop as a result of the interaction of these harmonic frequencies. For example, consider an FM receiver having an intermediate frequency of 9.1 Mc, and tuned to an 86 Mc station. The oscillator frequency is 86 + 9.1, or 95.1 Mc. It is possible that a strong 90.5 Mc signal picked up at the FM receiver input would develop at that point its second harmonic of 181.0 Mc. The oscillator second harmonic frequency is 95.1 x 2, or 190.2 Mc. The difference frequency is 190.2 minus 181.0, or 9.2 Mc. This difference frequency would appear in the output of the mixer stage and be accepted by the IF amplifiers tunned to 9.1 Mc. Thus the receiver output would contain the 86 Mc station and simultaneously the 90.5 Mc interfering signal.

Harmonics produced at the input may be reduced by increasing the selectivity of the tuned circuits and using variable mu tubes that do not overload easily. The production of harmonics by the local oscillator may be reduced by maintaining a satisfactorily high circuit Q and by reducing its loading.

In commercial FM IF amplifiers the bandpass is considerably greater than it is in AM IF amplifiers because of the greater frequency swing used in frequency modulation. An ideal frequency response curve is difficult to obtain economically. Therefore, a practical compromise that gives the necessary uniform gain and discrimination against adjacent channel frequencies is chosen.

The IF stage may be designed for FM only or for both AM and FM. An IF transformer designed for both AM and FM is shown in Figure 40-8.

In order to have the desired high L/C ratio for increased gain and increased bandwidth, permeability tuning is employed. Circuits C_1L_1 and L_2C_2 are tuned to the higher FM intermediate frequency, about 10 mc, and have greater bandpass, about 200 kc. Circuits C_3L_3 and L_4C_4 are tuned to the lower AM intermediate frequency, perhaps 455 kc, and the bandpass is lower, about 20 kc.

When the receiver is adjusted for FM reception, only the FM section of the IF transformer is effective in coupling signal voltage to the next tube. Capacitor C_3, having a low reactance to the higher FM signals, shunts the AM section of the transformer. Likewise, when the receiver is adjusted for AM reception, only the AM section of the IF transformer is effective in coupling signal voltage to the next

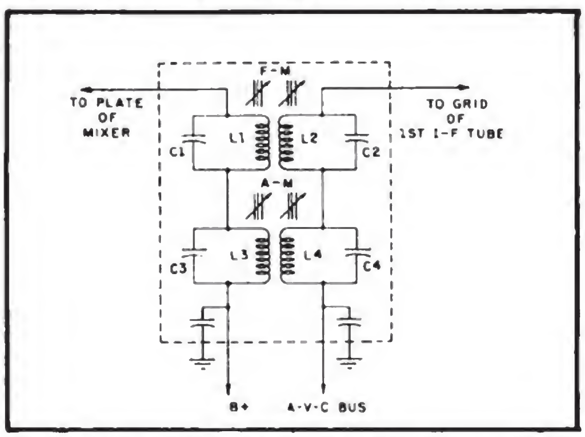

Figure 40-8 - IF transformers for AM and FM.

tube. In this case L_1 becomes an effective short circuit for the lower frequency AM signals. Usually the last IF stage is modified to operate as a limiter.

Q13. What factors determine the selection of the proper intermediate frequency?

40-10. Limiter

The limiter in an FM receiver removes amplitude variations and passes on to the discriminator an FM signal of constant amplitude.

As the FM signal leaves the transmitting antenna it is varying in frequency according to an audio modulating signal, but it has essentially a constant amplitude. As the signal travels between the transmitting and receiving antenna, however, natural and man made noises, or static disturbances, are combined with it to produce variations in the amplitude of the modulated signal. Other variations are caused by fading of the signal. Fading might be caused, for example, by movement of the ship carrying the transmitter or the receiver. Still other amplitude variations are introduced within the receiver itself because of a lack of uniform response of the tuned circuits.

All of these undesirable variations in the amplitude of the FM signal are amplified as the signal passes through the successive stages of the receiver up to the input of the limiter. This condition in which both frequency modulation (desired) and amplitude modulation (undesired) are present at the same time as shown in Figure 40-9A.

The character of the signal after leaving the limiter should be as indicated in Figure 40-9B, in which all amplitude variations have been removed, leaving a signal that varies only in frequency.

A grid leak bias limiter is shown in Figure 40-10. The tube is a sharp cutoff pentode operated with grid leak bias. Because the plate and screen voltages are purposely made low, plate current saturation as well as plate current cut-off, is produced readily by input signals having a magnitude of only a few volts.

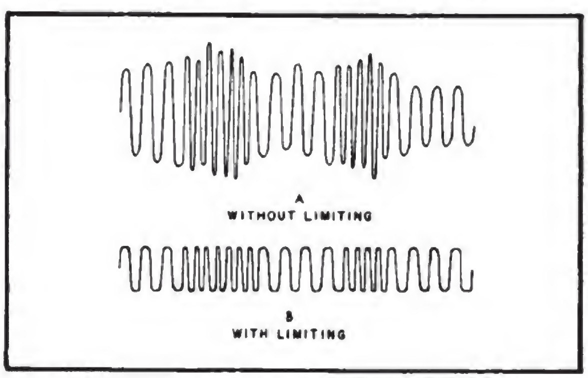

Figure 40-9 - FM signals.

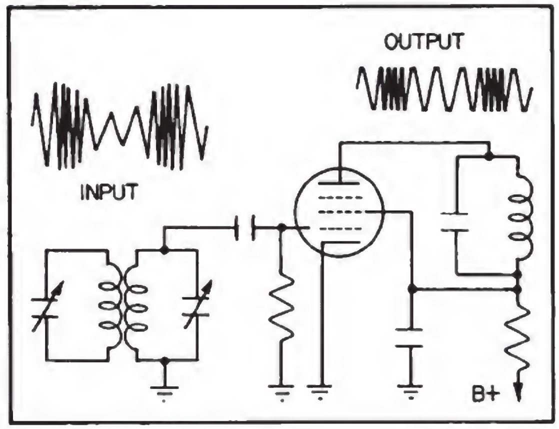

Figure 40-10 - Grid leak bias limiter.

The manner in which the limiter functions is illustrated by the e_c-i_b curve shown in Figure 40-11.

Grid leak bias is used so that with varying signal amplitudes, the bias can adjust itself automatically to a value that allows just the positive peaks of the signal to drive the grid positive and cause grid current to flow.

Suppose that a signal having a peak amplitude greater than the cutoff bias is impressed on the grid of the tube. A bias voltage having a magnitude approximately equal to the peak value of the signal will be developed. Accordingly, grid current will flow for a very small part of the positive half cycle at the peak of signal swing, as shown by the shaded area. Plate current flows for almost the entire positive half cycle. When the signal amplitude increases, a greater bias is developed, but the grid cutoff voltage remains the same and the

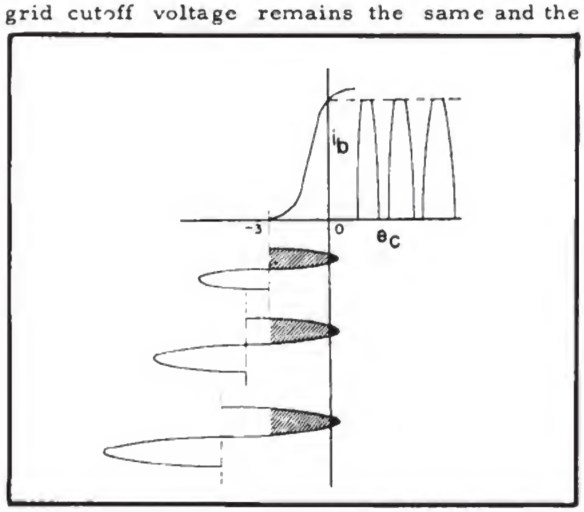

Figure 40-11 - Limiter e_c-i_b curve.

A13. Receiver gain, image response, selectivity, stability of local oscillator.

average plate current changes very little. Thus, the amount of output voltage in the limiter stage is approximately constant for all signals having an amplitude great enough to develop a grid leak bias voltage that is greater than the cutoff voltage. The frequency variations in the FM signal are maintained in the output because the plate current pulses are produced at the signal frequency and excite the plate tuned tank circuit which has a relatively low Q and a wide bandpass. Thus, because of the "flywheel" effect, a complete ac waveform is passed to the secondary of the discriminator transformer for each cycle of input signal.

When the peak amplitude of the grid signal is less than the cutoff voltage, the limiting action fails because the stage is practically a class A amplifier for such signals, and the average plate current varies as the grid leak bias changes with varying signal amplitudes. For this reason, the stages preceding the limiter must have sufficient gain to provide satisfactory limiting action on the weakest signal to be received.

Q14. Why are limiter stages operated with very low value of plate voltage?

Q15 Why is a sharp cutoff tube usually employed in a limiter stage instead of variable mu tube?

40-11. Demodulation of FM Waves

Another major difference between the AM receiver and the FM receiver is in the method used to detect the signal. The DETECTOR in an AM receiver interprets the AMPLITUDE VARIATIONS of the amplitude modulated RF energy in terms of the audio signal. In the FM receiver, the discriminator interprets the FREQUENCY VARIATIONS of the frequency modulated RF energy in terms of the audio signal.

In FM transmission the intelligence to be transmitted causes a variation in the instantaneous frequency of the carrier either above or below the center, or resting, frequency. The detecting device must therefore be so constructed that its output will vary linearly according to the instantaneous frequency of the incoming signal. Also, the detecting device must be insensitive to amplitude variation produced by interference or by receiver nonlinearities; thus a special limiting device, called a LIMITER, must precede the FM detector.

Several types of FM detectors have been developed and are in use, but perhaps two of the most common types are the Foster Seeley discriminator and the ratio detector.

The discriminator requires a limiter, which in turn requires considerable amplification ahead of its input.

An FM detector that would be insensitive to amplitude variations would eliminate the need for a limiter, and in addition one or more IF amplifier stages might be eliminated. Such an improved discriminator circuit that meets these requirements to a larger degree than the discriminator, is the RATIO DETECTOR. There are also various types of LOCKED OSCILLATOR FM detectors. The simplest type of detector is the SLOPE detector. Although it is rarely used, this type of FM detector will be considered first, because of its simplicity.

40-12. Slope Detector

Even an AM receiver may give a distorted reproduction of an FM signal under certain conditions of operation. When the carrier frequency of the FM signal falls on the sloping side of the RF response curve in an AM receiver, the frequency variations of the carrier signal are converted into equivalent amplitude variations. This conversion results from the unequal response above and below the carrier center frequency (point B), as shown in Figure 40-12.

Thus, when the incoming FM signal is less than the center frequency, for example, at point A, which is the minimum value, the output voltage is at a minimum in the negative direction. When the incoming signal swings to point C (the maximum value), the output voltage is maximum in the positive direction. The resultant AM signal may be coupled to the regular AM detector where the original audio voltage is recovered.

The obvious disadvantage of this type of detection is the nonlinearity of the response curve. At best, the most linear portion of the curve has a limited frequency range. Con-

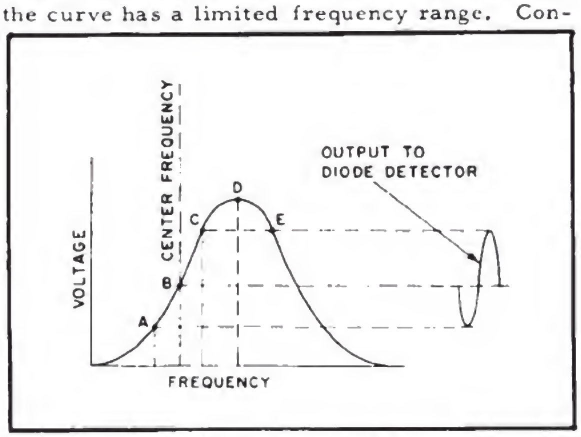

Figure 40-12 - Slope detection.

sequently the undistorted output voltage is low.

In Figure 40-12, if the center frequency falls at point D and the maximum frequency swings are between points C and E, there is no effective output signal voltage because the curve is relatively flat.

Q16. Why does distortion result when using slope detection?

40-13. Foster Seeley Discriminator Circuit

The detection of a change of frequency by the discriminator is accomplished by a transformer, which, due to special connections, distorts the frequency change into a voltage or varying amplitude change. After distortion, the voltage of varying amplitude is detected in the manner employed by any amplitude modulation detector. Figure 40-13 shows the circuit of a typical discriminator circuit.

The two coils L_1 and L_2 are the primary and the secondary, respectively, of the IF transformers. The secondary coils is tuned to the correct transmitter frequency by capacitor C. Tube V_1 and V_2 are diode detector tubes. The filters R_1C_1 and R_2C_2 serve the same purpose as the filters in the diode detector, that is, the removal of the RF component from the circuit.

The presence of L_3 and its connection to the primary at the top and to the secondary at the center presents an unusual feature in the discriminator circuit. L_3 is an RF choke which has a high reactance to the RF frequency. Since it is connected across the primary, the primary voltage appears across it at all times. The connection to each diode from this choke causes the primary voltage (E_p) to appear at the plates of the diode with the same phase shift in each case. This voltage causes currents to flow in the opposite directions in the resistors R_1 and R_2, resulting in zero output.

The phase relationship of the voltage across the RF choke to the voltage induced in the secondary is the key to the operation of the discriminator circuit. For this reason you will find it is desirable to analyze its operation in detail. To help make the analysis, Figure 40-14A shows the discriminator transformer, Figure 40-14B shows the vector relationships of current and voltage in the transformer, and Figure 40-14C shows the current and voltage waveshapes.

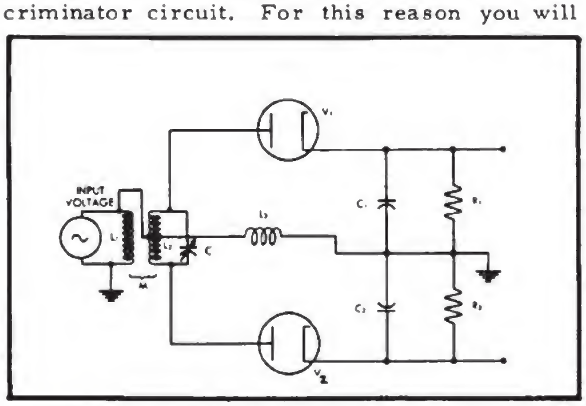

Figure 40-13 - Discriminator circuit.

At the desired frequency the transformer vector relationships are as follows: the primary voltage E_p is the reference vector. Since the mutual inductance (coupling) is small, the primary is inductive. The primary current I_p lags E_p by 90°. The magnetic field which affects the secondary (called the flux and shown by the symbol $\emptyset$) is in phase with the primary current. Due to normal transformer action, the voltage induced in the secondary is 90° behind the flux. This induced voltage, labeled E_i, is simulated by a generator in the equivalent circuit of the secondary, illustrated in Figure 40-15A. The generator is inserted at the center of the transformer secondary where the voltage across the secondary is divided between the two tubes. A current I_s flows around the loop composed of the secondary windings and the capacitor C. This loop is a series resonant circuit insofar as the generator sees it. Since the circuit is at resonance, the current I_s is in phase with the induced voltage.

Each half of the secondary has considerable inductive reactance at the RF frequency. Therefore, there is a voltage drop across this reactance due to the current I_s. These voltages are

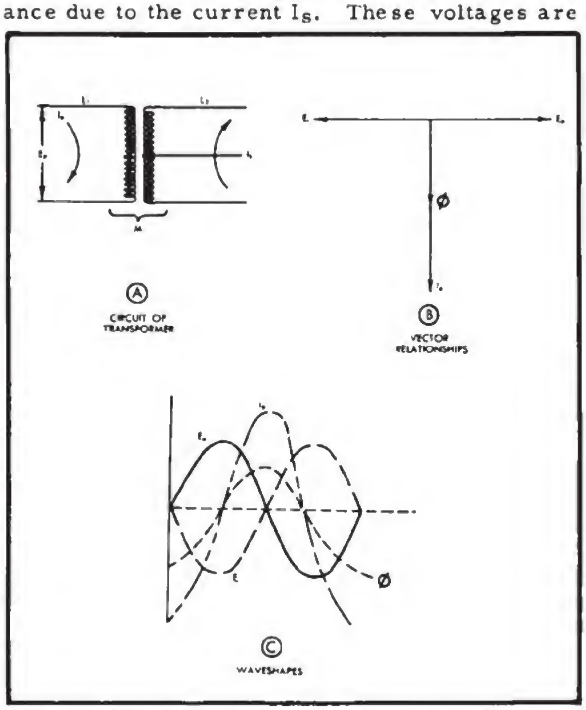

Figure 40-14 - Transformer vector relationships and waveshapes.

774-639 O - 65 - 13

A14. So that large values of positive grid voltage quickly drive the tube to saturation, and large negative values drive it to cutoff.

A15. So cutoff limiting can be achieved with low input signal strengths.

A16. Because of the non-linearity of the response curve.

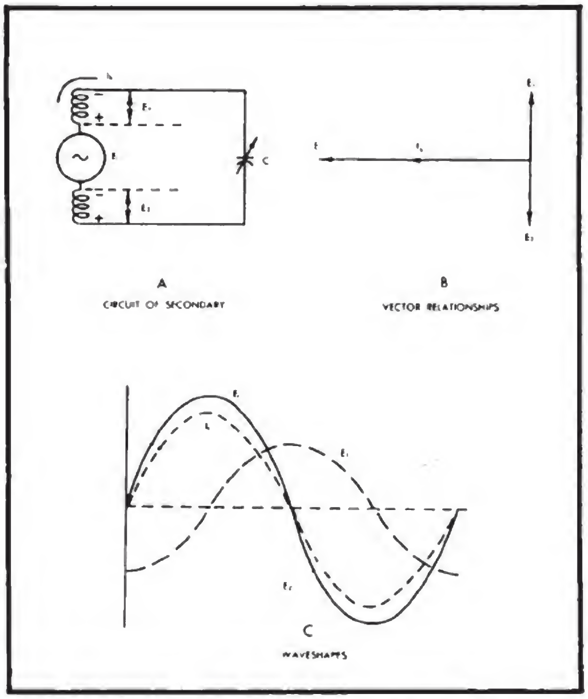

Figure 40-15 - Secondary voltages at resonance.

shown in Figure 40-15B as E_1 and E_2. Since voltage leads current by 90° in an inductor, E_1 and E_2 are 90° out of phase with I_s. The voltage E_2 actually leads I_s as it should, but the voltage E_1 is 180° out of phase; that is, its polarity is reversed, due to the manner in which it is connected to the tube. Current flows in the same direction through both coils, but the vector voltages are both measured with respect to the center. This causes opposite polarities to exist.

Referring to Figure 40-16A, the equivalent circuit of the discriminator in which the voltages are replaced by generators, note that the plate voltage on V_1 is the sum of the voltages E_p and E_1, and the plate voltage on V_2 is the sum of E_p and E_2. The vector diagram, Figure 40-16B, shows E_p and E_1 added vectorially to produce a vector sum E_{v1}. E_{v1} is the actual plate voltage on V_1. The two sine wave voltages E_p and E_1 are also shown. They add to produce a third sine

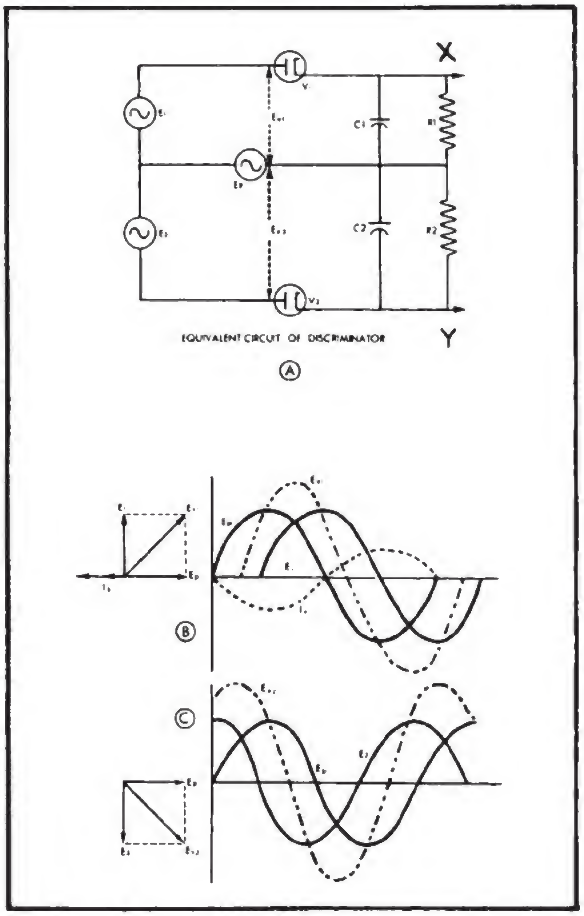

Figure 40-16 - Resultant plate voltages at resonance.

wave called E_{v1}. When plate current flows, C_1 charges to near the peak voltage of E_{v1}, producing a constant dc voltage almost equal to E_{v1} across resistor R_1. In the same manner, E_p and E_2 add as E_p and E_1 did previously resulting in E_{v2}, the plate voltage on the other diode. Consequently, C_2 charges to E_{v2} and a constant voltage almost equal to E_{v2} appears across R_2 as shown in Figure 40-16C. The vector and sine waves are shown.

Observe that the voltage across the filters R_1C_1 and R_2C_2 are dc voltages. Obviously, phase relationships cannot exist there. In this case only amplitude relationships are important. Since E_p is common to both and E_1 equals E_2, the vector sums of E_{v1} and E_{v2} are equal. The sine-wave illustrations show that the resultant of their sine-waves, although of different phase, have equal amplitudes. The dc voltages across each filter are alike, but have opposite polarities. Therefore, the sum of these voltages will be zero between X and Y. This fact indicates that the output at resonance is zero.

40-14. Output at Frequencies Below Resonance

When the frequency of an FM carrier wave is lower than the center or resonant frequency the output is no longer zero. Using the primary voltage as a reference, the voltage induced in the secondary is still 180° out of phase, as shown by the vector diagrams in Figure 40-17A. Since inductive reactance decreases as frequency decreases, while capacitive reactance increases, the secondary circuit of the transformer will be capacitive at frequencies below resonance. The current in a capacitive circuit leads the applied voltage. Here, the applied voltage is E_i, and I_s will lead. An arbitrary amount of lead is used in the vector diagram, Figure 40-17A. The voltage drop E_1 and E_2 will still be 90° out of phase with the current since this is not a function of frequency. Since under these conditions voltage E_1 is more than 90° away from E_p when added vectorially, their sum will be less than at resonance even through the magnitude of the components E_p and E_1 has not changed. At the same time, E_2 is less than 90° away from E_p, or they are more nearly in phase, so the magnitude of the resultant E_{v2} is greater than at resonance. Again, since the resistor voltages depend upon the length of the resultant vectors, there will be a higher voltage across R_2 and a lower voltage across R_1. Therefore, the output voltage will be equal to the R_2 voltage minus the R_1 voltage and the output point X will be negative with respect to Y which may be grounded. (Figure 40-16A)

The waveshape shows the same thing. I_s is shifted in phase; voltage E_1 is further out-of-phase with E_p, and E_2 is more nearly in phase with E_p. The sum of the more out-of-phase components is small; but the resultant waveshape of the two nearly in-phase components is large.

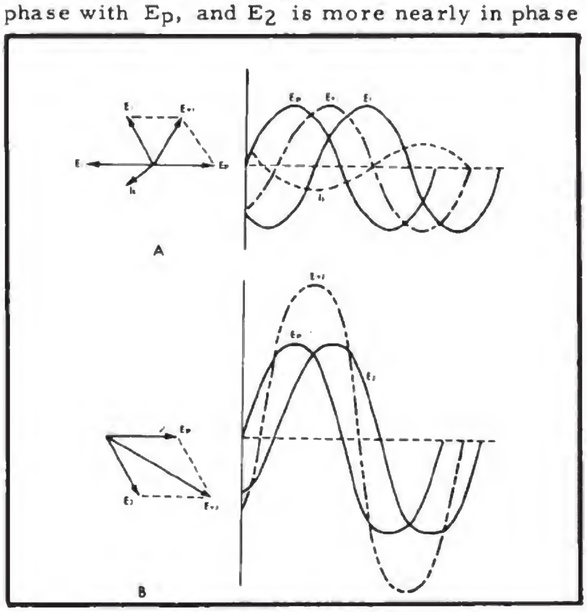

Figure 40-17 - Voltages at frequencies below resonance.

Studying the vector diagrams, you can see that the greater the deviation from resonance, the greater the dc output voltage.

40-15. Output at Frequencies Above Resonance

When the input frequency is greater than the center frequency, the circuit L_2C is inductive, and the current I_s lags the induced voltage. Figure 40-18 shows the vector relationship of voltages at frequencies above resonance. Note that at frequencies above resonance, E_1 is more nearly in-phase, and E_2 is more out-of-phase than at the center frequency. Further notice that E_{v1} has a greater magnitude than E_{v2}. Therefore, the dc voltage across R_1 is greater than the R_2 voltage, and consequently the output voltage is positive.

In summary, the application of a frequency modulated signal to the Foster Seeley discriminator results in the reproduction of the original modulating signal applied to the transmitter. When the frequency deviates below the resting frequency, the output voltage is negative at point X with respect to point Y. At the resting frequency, the output is zero and when the frequency deviates above the resting frequency, the output is positive. In the transmitter, the amount of frequency deviation was dependent upon the amplitude of the modulating signal. The amplitude of the audio signal developed in the output of the Foster Seeley discriminator is dependent on how far the frequency deviated from the center or resting frequency. 10% modulation would result in a small audio signal, whereas, 100% modulation would result in a large signal developed in the output circuit.

Q17. What would be the output of the discriminator (Figure 40-13) if the IF signal were to shift above its normal center value and stay there?

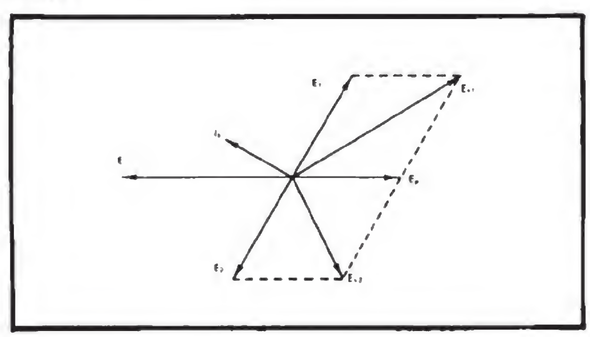

Figure 40-18 - Voltages at frequencies above resonance.

A17. The output from points X and Y would be a positive dc voltage. If the IF were modulated the audio signal would ride this dc reference voltage.

40-16. Ratio Detector

The Foster Seeley discriminator just described is insensitive to amplitude variations only at the resting frequency. When the frequency deviates, an unbalance occurs in the discriminator and produces a voltage output that is dependent upon the amount of frequency deviation and also upon the amplitude of the signal. Therefore, if the input signal to the discriminator is simultaneously varying in frequency and amplitude, the output from the discriminator will have the effects of both rather than of the frequency variation alone which will result in audio distortion. Therefore, a limiter is necessary to prevent such distortion. This disadvantage can be overcome by a ratio detector circuit which splits the rectified voltages in such a way that their ratio is directly proportional to the ratio of the applied IF voltages, which vary with frequency.

When the sum of the rectified voltages from the transformer is maintained at a constant value, the ratio between them must remain constant, and the individual rectified voltages also must be constant. Output, therefore, is independent of amplitude variations in the signal and no limiter is necessary. A simplified ratio detector circuit (Figure 40-19) shows both diodes connected so that their output adds, instead of subtracting as in the Foster Seeley discriminator. Capacitor C_L across the load resistors have a large value of capacitance and are charged by the output voltage of the rectifiers. This tends to make the total voltage across the load constant over the period of their respective RC time constants, since a large capacitor across the combined loads maintains an average signal amplitude that is adjusted automatically to the required operating level. The rectified output must not vary at audio frequency, and the time constant of the capacitor and its load resistor must be great enough to smooth out such changes. This time constant is approximately 0.5 seconds. The basic phase comparison circuit and the appropriate vector diagram of the ratio detector and the Foster Seeley discriminator are the same.

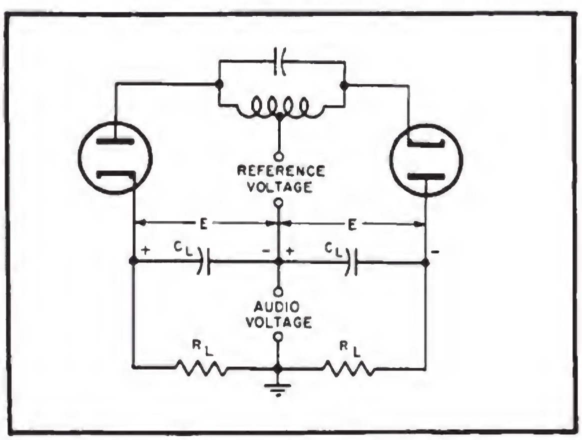

Figure 40-19 - Simplified ratio detector.

40-17. Practical Ratio Detector

In the circuit for a practical ratio detector (Figure 40-20) the voltages, E_1, E_2, and E_3, are obtained in the same way as in the Foster Seeley discriminator. Therefore, the applied voltage to the diodes is also the same. The diodes are connected in series, and the current through load resistor R_L is always in the same direction. Consequently, R_L acquires the polarity shown when the current flows from the plate of V_1 to the cathode of V_2. When an unmodulated signal is applied to the primary of the transformer, equal and opposite voltages E_2 and E_3 are developed across the secondary in respect to the center tap. These voltages are rectified by the diodes, with an output voltage across the load resistor, equal to their sum, or E_2 plus E_3, and the large capacitor, C_L, is charged to this constant voltage. The time constant of R_L and C_L is long compared with the lowest audio frequency.

Since the voltage across C_L is constant, the sum of the voltages across C_3 and C_4 also must remain fixed. When the carrier frequency shifts with modulation, however, the voltages across C_3 and C_4 change, but the sum of their voltages stays fixed at the amplitude of the charge on C_L. When the frequency decreases, C_4 acquires a greater charge than C_3; when the frequency increases, C_4 loses the charge to C_3. Therefore, the voltage between the center tap of the two capacitors and ground varies as the ratio of the voltages across C_3 and C_4, the ratio depending on the instantaneous frequency. A variable voltage whose amplitude depends on the frequency deviation of the carrier consequently can be applied to the audio output. As the rate of variation increases with frequency deviation, the voltage at the center tap changes frequency, producing a higher audio frequency. Any amplitude variation in the input signal to the transformers, no matter where the carrier is in its swing, also tends to change the voltage across C_3 and C_4. The voltage across the RC network, however, cannot change rapidly enough to follow the amplitude modulations, and the ratio of the voltage across C_3 and C_4 do not

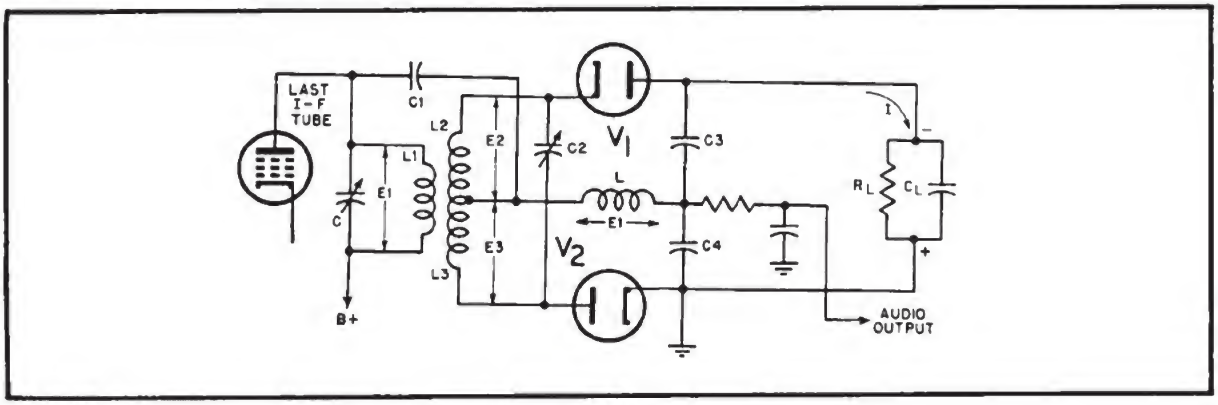

Figure 40-20 - Ratio detector.

change enough to produce an audio output.

40-18. Performance of Ratio Detector

The rectified voltage across the load circuit of the ratio detector adjusts itself to the amplitude of the input signal, and there is no minimum level where amplitude variation still can appear in the output. No matter how weak the signal is, the amplitude variations are removed to some extent by the constant charge on the capacitor. However, if signals of greater strength are tuned in, the charge on the capacitor is increased, and the total voltage across C_3 and C_4 is increased. Consequently, ratio detectors produce audio output that is proportional to the average strength of the received signal. Ratio detectors can operate with as little as 100 millivolts of input, which is much lower than that required for limiter saturation, and less IF gain is required. This receiver also is relatively quiet when no signal is received, since tube noise is not amplified as much.

Although the ratio detector requires fewer IF amplifier stages and minimum limiting, it presents alignment difficulties, and the signal may be distorted at high input voltages if some form of limiting is not applied.

Q18. How does the audio output voltage from the ratio detector compare to that of the Foster Seeley discriminator?

A18. Due to the series connection of the diodes the output from the ratio detector is equal to one half the output voltage of the Foster Seeley discriminator.

EXERCISE 40

1. What are the essential characteristics of an FM wave?
2. Describe what is meant by a mechanical modulator. Give an example of a mechanical modulator.
3. Define the following terms (a) resting frequency (b) frequency deviation (c) carrier swing (d) deviation ratio (e) rate of deviation.
4. What is the function of the reactance tube in the FM transmitter? Describe its input and output?
5. What do we mean when we refer to a sideband as being significant?
6. Explain the importance of the modulation index.
7. With a modulation index of 4 and a modulating frequency of 10 KC, compute the complete bandwidth required when a 95 Mc carrier is employed.
8. What are guardbands and why are they required in FM transmission?
9. With a carrier frequency of 90 Mc and the upper and lower sideband limits at $\pm$ 96 KC what modulation index was being employed at the transmitter if the frequency of the modulating signal were 12 kc?
10. If the frequency of the modulating signal is doubled, what happens to the FM wave? Describe in detail.
11. What is the relationship between the modulation index, the pairs of significant sidebands, and the power contained in the sideband pairs.
12. A twenty-five Mc carrier is frequency modulated 60% with a five kc signal. What is the bandwidth? Is it within the limits established by the FCC?
13. What is the significance of 100% modulation when applies to AM broadcasting? FM broadcasting?
14. What is the carrier frequency swing of an FM signal when the frequency deviation is (a) 40 KC, (b) 60 KC?
15. What is the relationship between an antenna used for FM and one used for AM?
16. A certain FM station operates on an assigned frequency of 105.5 Mc. (a) What is its wavelength in meters? (b) What is its wavelength in feet?
17. Compare the FM and AM systems with respect to their (a) operating frequencies (b) AF range (c) selectivity.
18. How do the requirements of the IF section of an FM receiver differ from those in an AM receiver?
19. Discuss the various factors which govern oscillator stability.
20. Why is oscillator stability so important in a high frequency receiver?
21. What is a limiter? Describe its input, output and operation.
22. Describe the phase relationships between primary and secondary of the Foster Seeley discriminator.
23. What is the relationship between the conduction of the diodes used in the Foster Seeley discriminator when the input is unmodulated?
24. What limitations of the Foster Seeley type of discriminator resulted in the development of a ratio detector?
25. How does the operation of a ratio detector differ from the operation of a Foster Seeley detector?

INDEX

U.S. GOVERNMENT PRINTING OFFICE : 1965 O - 774-639